Über den Charge Pumping-Effekt bei MOS-Transistoren

I. Untersuchung des Zusammenhangs zwischen Grenzschichtzuständen und Charge Pumping-Effekt bei MOS-Transistoren

II. Erweiterte Untersuchung des differentiellen Charge Pumping-Effekts bei MDS-Transistoren

INAUGURALDISSERTATION

zur Erlangung der Würde eines Doktors der Philosophie
vorgelegt der Philosophisch-Naturwissenschaftlichen
Fakultät der Universität Basel

von

JOHANNES GOLDER
aus Herdern (Thurgau)

Basel
Buchdruckerei Birkhäuser AG
1971

Über den Charge Pumping-Effekt bei MOS-Transistoren

I. Untersuchung des Zusammenhangs zwischen Grenzschichtzuständen und Charge Pumping-Effekt bei MOS-Transistoren

II. Erweiterte Untersuchung des differentiellen Charge Pumping-Effekts bei MDS-Transistoren

INAUGURALDISSERTATION

zur Erlangung der Würde eines Doktors der Philosophie
vorgelegt der Philosophisch-Naturwissenschaftlichen
Fakultät der Universität Basel

von

JOHANNES GOLDER
aus Herdern (Thurgau)

Basel
Buchdruckerei Birkhäuser AG
1972

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag der Herren PD Dr. W. Czaja und Prof. K. Alder

Basel, den 27. April 1971

Prof. Dr. H. P. Laubscher
Dekan

UNIV. LEIDEN BIBL.

Separatabdruck aus HPA *44*, 387–400 (1971) und HPA *44*, 941–958 (1971)

Untersuchung des Zusammenhangs zwischen Grenzschichtzuständen und Charge Pumping-Effekt bei MOS-Transistoren

von **J. Golder**

Institut für angewandte Physik der Universität Basel

(18. XI. 70)

Summary. As reported by Brugler and Jespers [1], periodical variations of the gate potential of enhancement-mode MOS-transistors cause a charge pumping effect correlated with the altering occupancy of states in the band gap. This paper extends the rather phenomenological analysis of ref. [1] and describes the effect more precisely. To this end, Shockley-Read-Hall statistics are, as usual, applied to the interface states. In contrast to the linear models treated by different authors, nonlinearities have to be considered in the analysis of the charge pumping. The theoretical result obtained for small periodical variations of the gate potential explains the experimental data well within the error limits. Thus it may be concluded, that under these conditions the interface states in p-type MOS-transistors are predominantly filled with electrons by the n-substrate and emptied by the p-regions (drain and source). This unidirectional electron transfer leads to the nonvanishing DC component in the substrate current observed in the charge pumping experiment. In principle, the energy distribution of the interface states might be determined by means of the charge pumping effect; but as a consequence of the nonlinear relations involved, this procedure becomes quite unpractical.

1. Einleitung

Die vorliegende Arbeit beschäftigt sich mit dem von Brugler und Jespers in [1] beschriebenen Charge Pumping-Effekt bei Enhancement-Mode MOS-Transistoren, der mit dem Umbesetzen von Grenzschichtzuständen und mit der Volumenrekombination zusammenhängt. In den üblichen MOS-Transistoren überwiegt der Grenzschichtanteil wesentlich. Die Oberflächenkonzentrationen p_s und n_s der Löcher respektive Elektronen ändern sich hauptsächlich durch Diffusion zwischen Oberfläche und Drain- sowie Source-Regionen respektive Substratvolumen. Aus diesem Grunde sind keine beliebig raschen Änderungen von p_s und n_s möglich.

Diese Arbeit soll in erster Linie abklären, inwieweit der Zusammenhang des Effekts mit dem Umbesetzen von Grenzschichtzuständen unmittelbare Rückschlüsse auf die Energieverteilung der Zustände erlaubt. Dazu genügt es, ein einfaches Modell der Oberfläche zu diskutieren. Bei mässig hohen Frequenzen darf in erster Näherung angenommen werden, dass die Oberflächenkonzentrationen p_s und n_s sich stets im Gleichgewicht befinden.

Figur 1 zeigt den Messaufbau. Drain (D) und Source (S) eines p-Enhancement-Mode MOS-Transistors sind geerdet. Das Substrat (B) ist an einen Elektrometerverstärker angeschlossen. Am Gate (G) wird eine Spannung U_G angelegt. Für die

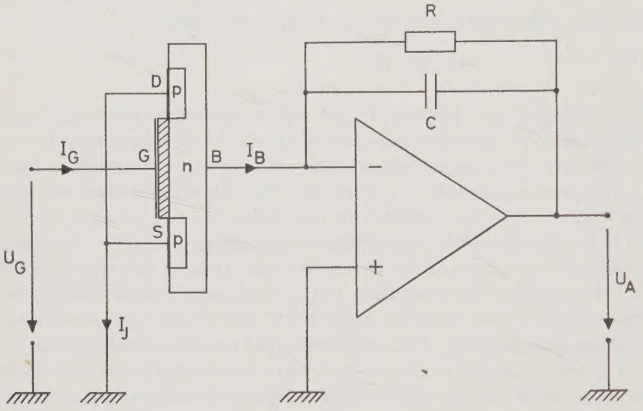
Mittelwerte der Ausgangsspannung U_A und des Eingangsstroms I_B gilt bei genügend starker Gegenkopplung:

$$\langle U_A \rangle = -R \langle I_B \rangle$$

und für kleine Änderungen δU_G um eine feste Vorspannung U_{GO} ergibt sich bei Frequenzen $f \gg 1/2\pi RC$ der folgende Zusammenhang:

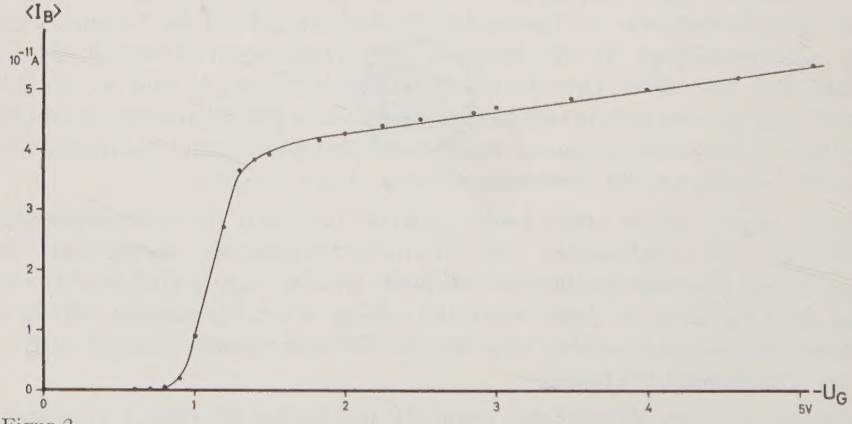
$$\delta U_A = -\delta U_G \frac{C_{GB}(U_{GO})}{C}$$

wobei $C_{GB}(U_{GO})$ die Teilkapazität zwischen Gate und Substrat bei geerdeten Source und Drain darstellt. Mit diesem Messaufbau lassen sich demzufolge Gleichströme und differentielle Kapazitäten messen.



Figur 1
Experimenteller Aufbau zur Messung von Gleichströmen und Kapazitäten.

Der Charge Pumping-Effekt äussert sich darin, dass eine periodische Änderung der Gatespannung einen von Null verschiedenen mittleren Substratstrom $\langle I_B \rangle$ erzeugt. Die Kurve in Figur 2 ergibt sich, wenn die Gatespannung dreieckförmig zwischen Erdpotential und $-U_G$ mit einer Frequenz von 1 kHz variiert wird. Die Angabe der Gatesignalform ist von Bedeutung, weil insbesondere die Flankensteilheiten einen starken Einfluss auf $\langle I_B \rangle$ ausüben.



Figur 2
Mittlerer Substratstrom $\langle I_B \rangle$ bei Großsignalaussteuerung. Die Gatespannung wird mit einer Frequenz von 1 kHz dreieckförmig zwischen Erdpotential und $-U_G$ variiert.

2. Allgemeine Grundlagen

Ein MOS-Transistor stellt zwischen Gate und den übrigen Anschlüssen im wesentlichen eine spannungsabhängige Kapazität dar. Bei grösseren Änderungen der Spannung U_G wird daher der Zusammenhang zwischen Ladung und Spannung nichtlinear. Bei periodischen Variationen treten im Gatestrom $I_G = I_B + I_J$ höhere Harmonische auf. Bei idealer Isolation gilt für den Mittelwert $\langle I_G \rangle$ unter allen Umständen:

$$\langle I_G \rangle = \langle I_B \rangle + \langle I_J \rangle = 0 \,. \tag{1}$$

Die experimentell festgestellte Abweichung von (1) um ungefähr $10^{-15} \ldots 10^{-14}$ A entspricht dem totalen Leckstrom bei der mittleren Gatespannung $\langle U_G \rangle$. Der dadurch entstehende Fehler darf bei sämtlichen in dieser Arbeit vorkommenden Messungen vernachlässigt werden.

Beim Charge Pumping misst man, wie aus Figur 1 hervorgeht, nur den Substratstrom I_B, für dessen linearen Mittelwert nach (1) gilt:

$$\langle I_B \rangle = -\langle I_J \rangle \,.$$

Damit $\langle I_B \rangle$ von Null verschieden sein kann, müssen folgende beiden Bedingungen zugleich erfüllt sein:

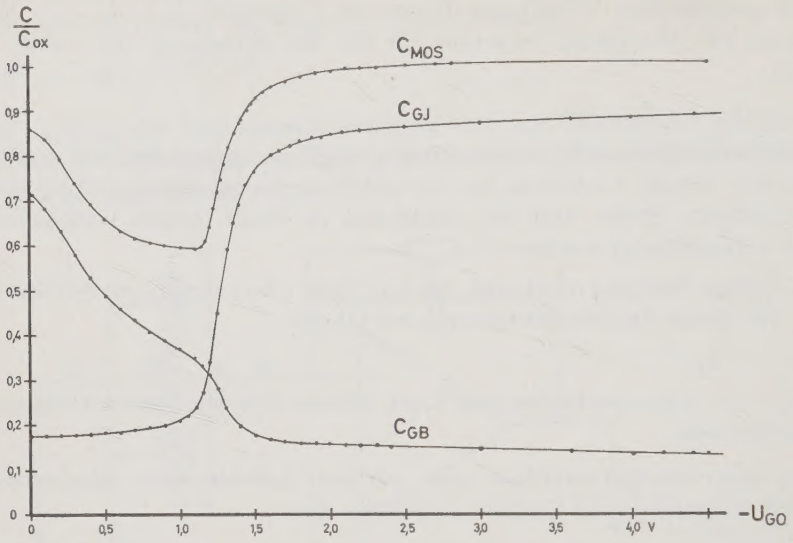
1. Neben dem Substratanschluss muss auf der Halbleiterseite mindestens ein weiterer stromführender Kontakt vorhanden sein.

2. Der Substratanschluss muss stärker an der Aufladung, die andern Kontakte müssen stärker an der Entladung der MOS-Kapazität beteiligt sein oder umgekehrt.

Zur ersten Bedingung ist folgendes zu sagen: Nach Jordan und Jordan [2] stellen die p-Gebiete (Source und Drain) Kontakte an das Valenzband des n-Substrats dar. Die Kopplung zwischen den p-Regionen und dem Substrat wird deshalb erst dann stark, wenn sich unter der Gateelektrode eine p-Inversionsschicht ausbildet. Solange die Gatespannung unter dem für Inversion notwendigen Schwellwert U_{TH} bleibt, fehlt praktisch der notwendige zweite Kontakt, und es wird $\langle I_B \rangle \approx \langle I_G \rangle = 0$. Das Experiment bestätigt diesen qualitativen Zusammenhang.

Der Substratanschluss bildet einen Kontakt an das Leitungsband des Substratvolumens. Damit lässt sich die Aufspaltung von I_G in die Anteile I_J und I_B noch genauer spezifieren: I_J wird von Löchern und I_B von Elektronen getragen. Zur Illustration dieses Sachverhaltes dient Figur 3, wo die Teilkapazitäten C_{GB} und C_{GJ} und die totale Kapazität C_{MOS} in Funktion der Gatevorspannung U_{GO} aufgetragen sind. C_{GJ} ist die Teilkapazität zwischen den p-Regionen und dem Gate bei geerdetem Substratanschluss. Nach Zaininger [3] entsteht die C_{MOS}-U_{GO}-Kurve wie folgt: Im Gebiet starker Akkumulation ist C_{MOS} ungefähr gleich der Oxidkapazität C_{ox}, da der grösste Teil der Ladung an der Halbleiteroberfläche sitzt. Im Übergangsgebiet werden die Majoritätsladungsträger (Elektronen im n-Substrat) von der Oberfläche verdrängt, und es resultiert eine Abnahme der Kapazität entsprechend derjenigen der Sperrschichtkapazität einer Schottky-Barriere. Darauf setzt die Bildung der Inversionsschicht an der Oberfläche ein, was erneut eine Zunahme der Kapazität bis zum

Wert C_{ox} bewirkt. Da jetzt die Inversionsschicht den Hauptteil der Ladung übernimmt, dringt das elektrische Feld nicht mehr wesentlich weiter ein. Aus Figur 3 wird klar, dass C_{GB} den Anteil der Substrat-Majoritätsladungsträger (Elektronen) und C_{GJ} denjenigen der Ladungsträger der Inversionsschicht (Löcher) darstellt.



Figur 3
Verlauf der Teilkapazitäten C_{GJ} und C_{GB} sowie der totalen MOS-Kapazität mit der Gatevorspannung $-U_{GO}$.
C_{MOS} Kapazität zwischen Gate und den übrigen Anschlüssen.
C_{GJ} Teilkapazität zwischen Gate und den p-Regionen bei geerdetem Substrat.
C_{GB} Teilkapazität zwischen Gate und Substrat bei geerdeten p-Regionen.

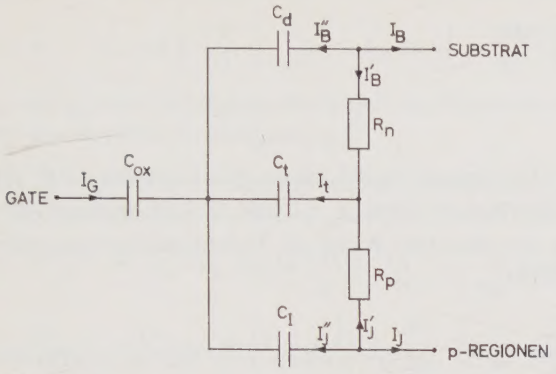
Zur zweiten Bedingung ist zu bemerken: Wenn man die Ladung Q_0 einer MOS-Kapazität um ΔQ erhöht, wird ΔQ von den p-Regionen und dem Substrat zusammen geliefert. Der Substratanteil sei Δq_1. Bringt man die MOS-Kapazität in den Ausgangszustand mit der Ladung Q_0 zurück, so wird ΔQ an die p-Regionen und an das Substrat abgegeben. Der Substratanteil sei nun Δq_2. Ein von Null verschiedener mittlerer Substratstrom entsteht bei periodischer Wiederholung nur dann, wenn gilt:

$$\Delta q_1 - \Delta q_2 = \Delta q \neq 0 .$$

Die Ladung Δq wird von den p-Regionen pro Zyklus an das Substrat geliefert. Da die p-Regionen Kontakte an das Valenzband, das Substrat an das Leitungsband darstellen, kommen für den Ladungstransfer nur Mechanismen in Frage, welche beide Bänder betreffen wie Rekombination und Umbesetzen von Zuständen in der verbotenen Zone. In der Folge wird nur das Umbesetzen von Grenzschichtzuständen berücksichtigt.

Für kleine Variationen um eine Gleichgewichtslage lässt sich das MOS-System in Anlehnung an Jordan und Jordan [2], Nicollian und Goetzberger [4] und Lehovec und Slobodskoy [5] durch das Modell in Figur 4 darstellen. Die Diffusion durch die PN-Übergänge und vom Substratvolumen durch die Depletion-Zone an die Oberfläche

sind in diesem Modell nicht enthalten, sodass die Dichten der Löcher p_s und Elektronen n_s an der Oberfläche jeder Änderung der Bandrandverbiegung ohne Verzögerung folgen können. Damit befinden sich die Ladungsträger in den Bändern stets im Gleichgewicht, und es gilt für die Quasiferminiveaux: $F_n = F_p$.



Figur 4
Vereinfachtes Ersatzschema für das MOS-System.
C_{ox} Oxidkapazität,
C_d Depletionkapazität,
C_I Inversionskapazität,
C_t Kapazität der Grenzschichtzustände,
R_n Ersatzwiderstand für den Ladungsaustausch mit dem Leitungsband,
R_p Ersatzwiderstand für den Ladungsaustausch mit dem Valenzband.

Das lineare Modell von Figur 4 gilt nur für Variationen der Brandrandverbiegung, welche im Vergleich zu kT sehr klein sind. Die beim Charge Pumping verwendeten Amplituden erzeugen jedoch bedeutend grössere Verbiegungen, und demzufolge sind die Vernachlässigungen, die zum linearen Modell führen, nicht mehr zulässig.

3. Nichtlineare Theorie und experimentelle Ergebnisse

Das Ziel dieses Kapitels ist die Berechnung des mittleren Substratstroms $\langle I_B \rangle$ und der Vergleich von theoretischen und experimentellen Ergebnissen. Da nur der Fall relativ kleiner Variationen $\Delta\psi_s$ um eine Gleichgewichtslage ψ_{so} im Inversionsgebiet eingehend diskutiert wird, kann man der Rechnung nach Nicollian und Goetzberger [4] ein diskretes Niveau E_t der Grenzschichtzustände zu Grund legen. Zur Vereinfachung nimmt man an, dass die Zustände homogen über die gesamte Gatefläche verteilt sind.

Aufbauend auf die Shockley-Read-Hall-Statistik wird zunächst ein allgemeiner Ansatz für die Umbesetzung der Grenzschichtzustände angegeben. Dieser führt auf eine in der Variation $\Delta\psi_s$ nichtlineare Differentialgleichung.

Der zweite Abschnitt umfasst die approximative Lösung und die Berechnung des mittleren Substratstroms $\langle I_B \rangle$ für kleine Änderungen $\Delta\psi_s$ der Bandrandverbiegung.

Im letzten Abschnitt werden die theoretischen und die experimentellen Ergebnisse verglichen.

3.1. *Allgemeiner Ansatz*

Die Fermistatistik schreibt für jede Bandrandverbiegung ψ_s die Besetzung n_t der Grenzschichtzustände vor. Mit der Anzahl N_t dieser Zustände pro Flächeneinheit gilt:

$$n_t = N_t f_t \tag{2}$$

wobei f_t die Fermifunktion bedeutet:

$$f_t = \frac{1}{1 + \exp\left(\frac{E_t - F_t}{kT}\right)}$$

E_t ist die Energie der Grenzschichtzustände, welche bezüglich der Energie E_i (Mitte der verbotenen Zone an der Oberfläche) durch $E_t = \Delta E_t + E_i$ festgelegt sei. Die Bandrandverbiegung ψ_s wird in der üblichen Weise als Potentialdifferenz zwischen Oberfläche und Volumen eingeführt:

$$q\,\psi_s = E_{iB} - E_i\,.$$

E_{iB} bezeichnet die Energie in der Mitte der verbotenen Zone im Substratvolumen. Für E_t ergibt sich also $E_t = \Delta E_t + E_{iB} - q\,\psi_s$, und durch Einsetzen von E_t folgt die Abhängigkeit der Fermifunktion f_t von der Bandrandverbiegung : $f_t = f_t(\Delta E_t, \psi_s)$.

Für die zeitliche Änderung der Besetzung n_t gilt mit der Nettoeinfangsrate U_{cn} für Elektronen aus dem Leitungsband und mit der Nettoeinfangsrate U_{cp} für Löcher aus dem Valenzband:

$$\frac{dn_t}{dt} = U_{cn} - U_{cp}\,. \tag{3}$$

Die in den Grenzschichtzuständen sitzende Ladung Q_t pro Flächeneinheit berechnet sich zu $Q_t = -q\,n_t$, und mit der Gatefläche A wird:

$$A\,\frac{dQ_t}{dt} = -q\,A\,U_{cn} + q\,A\,U_{cp}\,.$$

$A(dQ_t/dt)$ ist gleich dem gesamten Ladestrom I_t der Grenzschichtkapazität C_t im Modell von Figur 4, $-q\,A\,U_{cn}$ und $q\,A\,U_{cp}$ sind die Anteile, welche das Leitungsband respektive das Valenzband beitragen:

$$A\frac{dQ_t}{dt} = I_t = -q\,A\,N_t\frac{df_t}{dt} \qquad I'_B = -q\,A\,U_{cn} \qquad I'_J = q\,A\,U_{cp}\,. \tag{4) (4a) (4b}$$

Nach Shockley und Read [6] gilt für U_{cn} und U_{cp}:

$$U_{cn} = N_t\langle c_n\rangle\left[(1 - f_t)\,n_s - f_t\,n_1\right],$$

$$U_{cp} = N_t\langle c_p\rangle\left[f_t\,p_s - (1 - f_t)\,p_1\right].$$

$N_t\langle c_n\rangle$ Wahrscheinlichkeit pro Zeiteinheit für den Einfang eines Elektrons aus dem Leitungsband, wenn alle Grenzschichtzustände leer sind.

$N_t\langle c_p\rangle$ Wahrscheinlichkeit pro Zeiteinheit für den Einfang eines Lochs aus dem Valenzband, wenn alle Grenzschichtzustände gefüllt sind.

n_1, p_1 Dichte der Elektronen beziehungsweise der Löcher, wenn das Ferminiveau mit der Energie E_t zusammenfällt.

$$n_1 = n_i \exp\left(\frac{\Delta E_t}{kT}\right) \qquad p_1 = n_i \exp\left(-\frac{\Delta E_t}{kT}\right) \qquad n_1 p_1 = n_i^2 .$$

n_s, p_s Dichte der Elektronen beziehungsweise der Löcher an der Oberfläche.

Für $F_p = F_n = F$ lassen sich n_s und p_s nach Goetzberger und Sze [7] schreiben:

$$n_s = n_B \exp\left(\frac{q\psi_s}{kT}\right) \qquad p_s = p_B \exp\left(-\frac{q\psi_s}{kT}\right) \qquad n_s p_s = n_B p_B = n_i^2$$

n_B und p_B sind die Konzentrationen im Substratvolumen und als solche unabhängig von der Bandrandverbiegung ψ_s.

Setzt man n_s und p_s in die Beziehungen für U_{cn} und U_{cp} ein, so resultiert unter Berücksichtigung von $n_B p_B = n_i^2$ sowie $n_1 p_1 = n_i^2$:

$$U_{cn} = N_t \langle c_n \rangle n_B \exp(u_s) \left\{1 - f_t\left[1 + \frac{n_1}{n_B} \exp(-u_s)\right]\right\} \tag{5a}$$

$$U_{cp} = -N_t \langle c_p \rangle p_1 \left\{1 - f_t\left[1 + \frac{n_1}{n_B} \exp(-u_s)\right]\right\} \tag{5b}$$

wo $u_s = (q\psi_s/kT)$ die normierte Bandrandverbiegung bedeutet.

Aus (2) und (3) folgt:

$$\frac{df_t}{dt} = \frac{1}{N_t}(U_{cn} - U_{cp})$$

und mit (5a) und (5b):

$$\frac{df_t}{dt} = \left[\langle c_n \rangle n_B \exp(u_s) + \langle c_p \rangle p_1\right] \left\{1 - f_t\left[1 + \frac{n_1}{n_B} \exp(-u_s)\right]\right\} \tag{6}$$

Einsetzen von (5a) in (4a) ergibt:

$$I_B' = -q A N_t \langle c_n \rangle n_B \exp(u_s) \left\{1 - f_t\left[1 + \frac{n_1}{n_B} \exp(-u_s)\right]\right\} \tag{7}$$

Die Beziehung (6) stellt eine lineare Differentialgleichung für die Fermifunktion f_t dar.

Bei vorgegebener periodischer Störung $u_s = u_s(t)$ folgt $f_t = f_t(t)$ aus (6), und damit lässt sich der Mittelwert $\langle I_B' \rangle$ berechnen. Nach Figur 4 gilt für die Ströme I_B, I_B' und I_B'':

$$I_B = I_B' + I_B'' \qquad I_B'' = 0 .$$

Wegen $\langle I_B'' \rangle = 0$ resultiert für den Mittelwert $\langle I_B \rangle$ des Substratstroms:

$$\langle I_B \rangle = -\langle I_B' \rangle . \tag{8}$$

3.2. *Der Fall kleiner Variationen* $\Delta\psi_s$

Setzt sich die Bandrandverbiegung aus einem konstanten Anteil ψ_{so} und einem zeitlich periodischen Anteil $\Delta\psi_s \sin\omega t = (kT/q)\, a \sin\omega t$ zusammen, so wird mit $n_{so} = n_B \exp(\psi_{so})$ und $p_{so} = p_B \exp(-\psi_{so})$ aus der Differentialgleichung (6):

$$\frac{df_t}{dt} = \left[\langle c_n \rangle n_{so} \exp(a \sin\omega t) + \langle c_p \rangle p_1\right] \left\{1 - f_t\left[1 + \frac{n_1}{n_{so}} \exp(-a \sin\omega t)\right]\right\} \tag{9}$$

Für I'_B ergibt sich aus (7):

$$I'_B = -q\,A\,N_t\,\langle c_n\rangle\,n_{so}\exp(a\sin\omega t)\left\{1-f_t\left[1+\frac{n_1}{n_{so}}\exp(-a\sin\omega t)\right]\right\} \tag{10}$$

Wie im Anhang gezeigt wird, berechnet sich der mittlere Substratstrom $\langle I_B\rangle$ zu:

$$\langle I_B\rangle = q\,A\,N_t\,\langle c_n\rangle\,\langle c_p\rangle\,n_i^2\,\frac{\tau}{2}\left(\frac{q\,\Delta\psi_s}{kT}\right)^2\frac{\omega^2\tau^2}{1+\omega^2\tau^2} \tag{13}$$

Der reziproke Wert der Zeitkonstanten τ erfüllt dabei die folgende Beziehung:

$$\frac{1}{\tau} = \frac{\langle c_n\rangle\,n_{so}}{f_{to}} + \frac{\langle c_p\rangle\,p_{so}}{1-f_{to}} = \frac{1}{\tau_{ns}} + \frac{1}{\tau_{ps}}$$

τ_{ns} und τ_{ps} sind identisch mit den von Nicollian und Goetzberger [4] definierten Zeitkonstanten für Übergänge von Majoritätsladungsträgern respektive Minoritätsladungsträgern. Mit τ_{ns} und τ_{ps} lässt sich (13) schreiben als:

$$\langle I_B\rangle = q\,A\,N_t\frac{f_{to}(1-f_{to})}{\tau_{ns}+\tau_{ps}}\,\frac{1}{2}\left(\frac{q\,\Delta\psi_s}{kT}\right)^2\frac{\omega^2\tau^2}{1+\omega^2\tau^2} \tag{14}$$

Die Theorie liefert $\langle I_B\rangle$ als Funktion von $\Delta\psi_s$; experimentell bestimmt man $\langle I_B\rangle$ in Abhängigkeit von ΔU_G. Zum Vergleich der theoretischen und der experimentellen Ergebnisse muss also der Zusammenhang zwischen $\Delta\psi_s$ und ΔU_G ermittelt werden.

Für eine Ladungsänderung ΔQ_G auf der Gateelektrode gilt:

$$\Delta Q_G = \int_{U_{GO}}^{U_{GO}+\Delta U_G} C_{MOS}\,dU_{GO}\,.$$

Die Spannung U_{ox} über der Oxidschicht hat den Wert: $U_{ox} = U_G - \psi_s$, und es wird $\Delta Q_G = C_{ox}\,\Delta U_{ox} = C_{ox}\,(\Delta U_G - \Delta\psi_s)$, und schliesslich resultiert in Übereinstimmung mit Berglund [8]:

$$\Delta\psi_s = \Delta U_G - \int_{U_{GO}}^{U_{GO}+\Delta U_G}\frac{C_{MOS}}{C_{ox}}\,dU_{GO}$$

Das Integral lässt sich aus der gemessenen $C_{MOS}(U_{GO})$-Kurve durch graphische Integration erhalten.

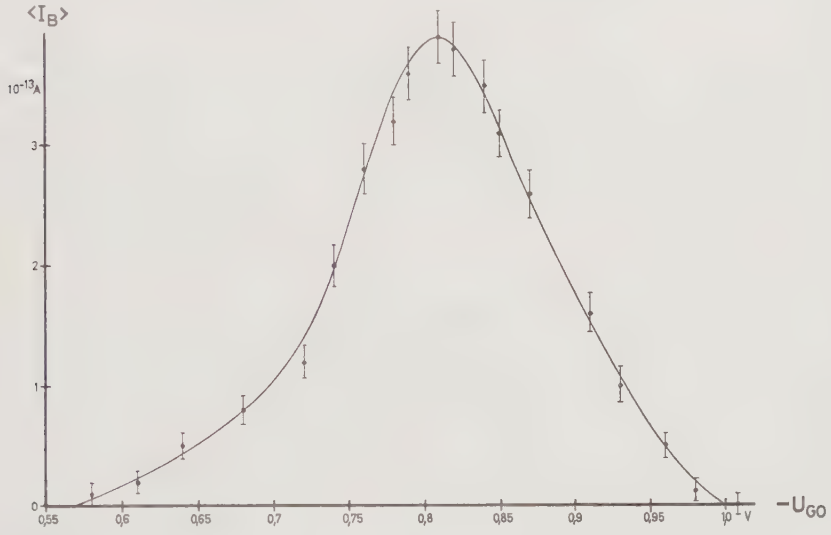
3.3. *Vergleich zwischen Näherungstheorie und Experiment*

Die ideale Kopplung zwischen p-Regionen und Oberfläche, die für die Herleitung von (14) vorausgesetzt wurde, besteht nur, wenn eine Inversionsschicht vorhanden ist. Für das Akkumulations- sowie das Depletiongebiet verschwindet der gemittelte Substratstrom $\langle I_B\rangle$, denn die Kopplung wird mit abnehmender Löcherkonzentration p_s immer schwächer, und der für $\langle I_B\rangle \neq 0$ notwendige zweite Kontakt geht schliesslich verloren. Der Gültigkeitsbereich der Beziehung (14) erstreckt sich demzufolge ausschliesslich auf das Inversionsgebiet. Weiter wurde bei der Herleitung von (14) die Diffusion der Ladungsträger vom Substratvolumen und von den p-Regionen nach der Oberfläche nicht berücksichtigt. Die Diffusion bewirkt eine Modifikation der

Frequenzabhängigkeit von $\langle I_B \rangle$ gegen hohe Frequenzen, da n_s und p_s den Variationen des Oberflächenpotentials ψ_s nicht beliebig rasch zu folgen vermögen.

Zu den experimentellen Ergebnissen ist folgendes zu bemerken: Aus Figur 3 bestimmt sich der minimale Wert von C_{MOS}/C_{ox} zu 0,6, so dass man für $\Delta\psi_s$ max $\approx$ 0,4 ΔU_G erhält. Die approximative Theorie verlangt $a = (q\,\Delta\psi_s/kT) \ll 1$, also ΔU_Geff $\ll$ (25,6 mV/1,4 · 0,4) $\approx$ 45 mV bei Raumtemperatur.

Figur 5 gibt eine Übersicht über den Verlauf des mittleren Substratstroms $\langle I_B \rangle$ als Funktion der Gatevorspannung U_{GO} (ΔU_Geff = 10 mV, f = 10 kHz). Zur Überprüfung der Theorie für kleine Signale sind zwei Messreihen aufgenommen worden: die eine für den Zusammenhang zwischen $\langle I_B \rangle$ und der Amplitude ΔU_G und die andere für die Abhängigkeit von $\langle I_B \rangle$ von der Frequenz f.

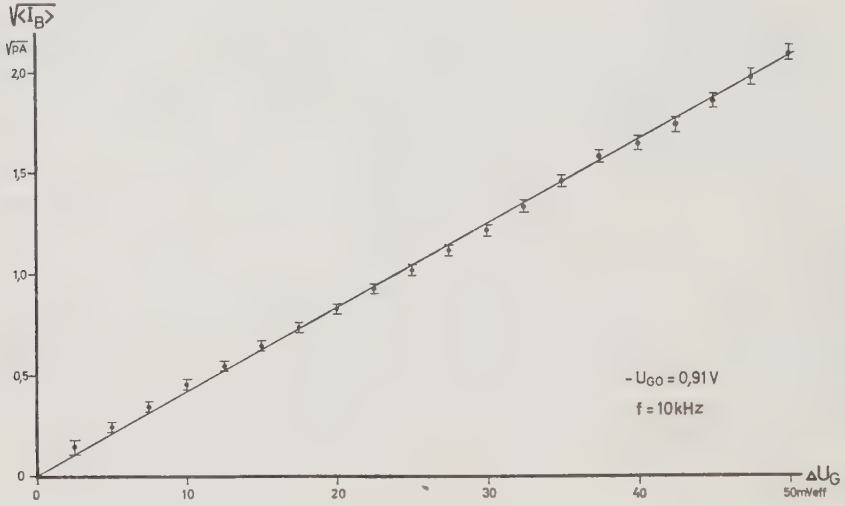


Figur 5
Mittlerer Substratstrom $\langle I_B \rangle$ bei sinusförmiger Kleinsignalaussteuerung mit einer Frequenz von 10 kHz. Die Signalgrösse ΔU_G beträgt 10 mVeff.

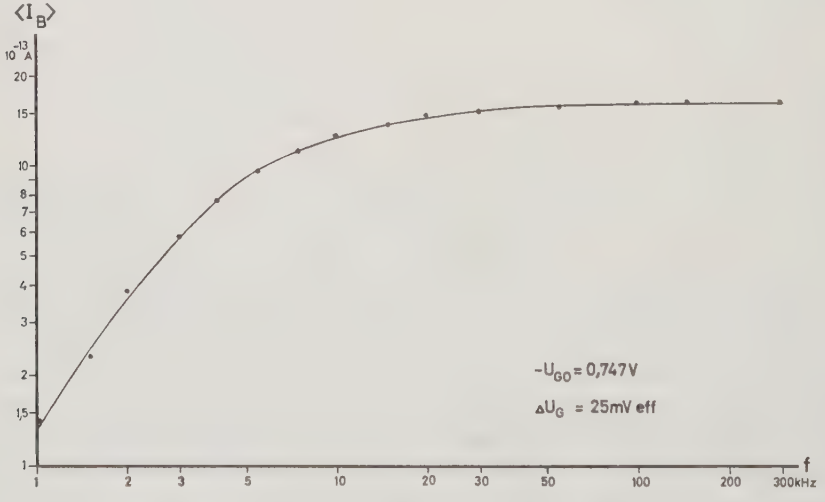
Figur 6 zeigt das Ergebnis des ersten Experiments. Bei konstanter Gatevorspannung $U_{GO} = -0{,}91$ V und konstanter Frequenz f = 10 kHz ergibt sich innerhalb der Fehlergrenzen die theoretisch vorausgesagte Proportionalität zwischen $\langle I_B \rangle$ und ΔU_G^2 (beziehungsweise $\Delta\psi_s^2$). Beachtenswert ist, dass die experimentell festgestellte Proportionalität nicht nur für ΔU_Geff $\ll$ 45 mV, sondern bis ΔU_Geff $\approx$ 50 mV gilt.

Figur 7 veranschaulicht die Frequenzabhängigkeit von $\langle I_B \rangle$ bei konstanter Amplitude ΔU_G bei einer festen Gatevorspannung U_{GO}. Die verwendete Apparatur mit einer Auflösung von 10^{-14} A erlaubte es nicht, diese Messung unter Einhaltung von ΔU_Geff $\ll$ 45 mV vernünftig genau durchzuführen. Daher wurde eine grössere Amplitude gewählt, welche jedoch immer noch im Bereich der Proportionalität zwischen $\langle I_B \rangle$ und ΔU_G^2 lag. Ein Kontrollversuch mit ΔU_Geff = 10 mV ergab unter Berücksichtigung der reduzierten Genauigkeit des Verstärkers dieselben Ergebnisse.

Aus dieser doppelt logarithmischen Darstellung von $\langle I_B \rangle$ als Funktion der Frequenz lässt sich die Zeitkonstante τ als Schnittpunktsabszisse der Kurvenasymptoten für den angegebenen Arbeitspunkt ($U_{GO} = -0{,}747$ V) bestimmen. τ ist die charakteristische Zeitkonstante für die Grenzschichtzustände, welche sich in der Nähe des Fermiveaus befinden. Beim Charge Pumping mit kleinen Spannungsänderungen ΔU_G um eine feste Vorspannung U_{GO} tragen nur Grenzschichtzustände in einem engen Energiebereich um die Gleichgewichtslage bei.



Figur 6
Abhängigkeit des mittleren Substratstroms $\langle I_B \rangle$ von der Amplitude ΔU_G bei konstanter Vorspannung $-U_{GO}$ und konstanter Frequenz.



Figur 7
Abhängigkeit des mittleren Substratstroms $\langle I_B \rangle$ von der Frequenz f bei konstanter Amplitude ΔU_G und konstanter Vorspannung $-U_{GO}$.

4. Diskussion

Aus der Übereinstimmung der Theorie mit den in den Figuren 6 und 7 wiedergegebenen experimentellen Ergebnissen lassen sich zusammenfassend folgende Schlüsse ziehen:

Bei den untersuchten p-Enhancement-Mode MOS-Transistoren entsteht der mittlere Substratstrom dadurch, dass die Grenzschichtzustände vorwiegend aus dem n-Substrat mit Elektronen aufgefüllt und von den p-Regionen (Drain und Source) entleert werden. Die Ursache für die asymmetrischen Umladevorgänge liegt in den verschieden nichtlinearen Zusammenhängen zwischen der Bandrandverbiegung und der Einfangsrate für Elektronen einerseits sowie der Einfangsrate für Löcher andererseits.

Bei kleinen Variationen $\Delta\psi_s$ um eine Gleichgewichtslage ψ_{s0} sind nur Zustände in einem schmalen Energieintervall um das Ferminiveau herum beteiligt. Daher ist es prinzipiell möglich, aus differentiellen Charge Pumping-Messungen die Energieverteilung der Grenzschichtzustände zu bestimmen. Die in der Theorie auftretenden Nichtlinearitäten haben jedoch zur Folge, dass die Auswertung der experimentellen Ergebnisse umständlicher wird als bei linearen Methoden nach Nicollian und Goetzberger [4] oder Berglund [8].

Die lineare Theorie von Nicollian und Goetzberger [4] beschreibt nur Umladevorgänge, bei welchen kein Nettoladungstransfer zwischen Valenzband und Leitungsband stattfindet. Das in dieser Arbeit behandelte Charge Pumping mit kleinen Signalen erlaubt im Quasigleichgewichtszustand ohne grossen Aufwand eine Untersuchung der Umladevorgänge mit Nettoladungstransfer zwischen den beiden Bändern. Da alle anderen Umladevorgänge keinen Beitrag zum gemessenen mittleren Substratstrom ergeben, wird eine äusserst gute Trennung erreicht. Eine Erweiterung der vorliegenden Theorie auf beliebig grosse und rasche Änderungen der Bandrandverbiegung würde es ermöglichen, die Analyse über den Quasigleichgewichtszustand hinaus auf allgemeine Nichtgleichgewichtszustände auszudehnen. Dazu müsste insbesondere die Diffusion vom Substratvolumen und von den p-Regionen nach der Oberfläche mitberücksichtigt werden.

Der theoretische Teil dieser Arbeit (Kapitel 3) basiert auf dem Modell von Figur 4, in welchem die Einflüsse der Diffusion nicht enthalten sind. Deshalb erklärt die Gleichung (14) einerseits die Abhängigkeit von der Gleichgewichtslage nach schwächerer Inversion und andererseits den Frequenzgang nach hohen Frequenzen hin nicht vollständig.

Brugler und Jespers [1] erwähnen, dass der Charge Pumping-Effekt bei der Anwendung von MOS-Transistoren als Analogschalter einen störenden Offsetstrom erzeugt. Dies ist von praktischer Bedeutung, da MOS-Transistoren oft als Schalter für kleine Signale verwendet werden. Eine auf beliebig grosse Änderungen der Bandrandverbiegung ausgedehnte theoretische Analyse liesse die Diskussion dieses technisch wichtigen Problems zu.

Anhang

Dieser Anhang umfasst die Herleitung des mittleren Substratstroms $\langle I_B \rangle$ für den Fall kleiner Variationen ausgehend von den Gleichungen (9) und (10):

$$\frac{df_t}{dt} = [\langle c_n \rangle n_{so} \exp(a \sin\omega t) + \langle c_p \rangle p_1] \left\{1 - f_t\left[1 + \frac{n_1}{n_{so}} \exp(-a \sin\omega t)\right]\right\} \tag{9}$$

$$I'_B = -q\, A\, N_t \langle c_n \rangle n_{so} \exp(a \sin\omega t) \left\{1 - f_t\left[1 + \frac{n_1}{n_{so}} \exp(-a \sin\omega t)\right]\right\} \tag{10}$$

Nach Einsetzen von (4) und (10) in (9) folgt:

$$-\frac{I_t}{q\, A\, N_t} - -\frac{I'_B}{q\, A\, N_t} + \langle c_p \rangle p_1 \left\{1 - f_t\left[1 + \frac{n_1}{n_{so}} \exp(-a \sin\omega t)\right]\right\}$$

und wegen $\langle I_t \rangle = 0$ resultiert für $\langle I_B \rangle$ mit (8):

$$\langle I_B \rangle = -\langle I'_B \rangle = -q\, A\, N_t \langle c_p \rangle p_1 \frac{1}{\omega T} \int_0^{0+2\pi} \left\{1 - f_t\left[1 + \frac{n_1}{n_{so}} \exp(-a \sin\omega t)\right]\right\} d(\omega t)\,.$$

Für f_t wird eine Fourierreihe angesetzt:

$$f_t = \alpha_0 + \sum_{\mu=1}^{\infty} \alpha_\mu \cos(\mu\omega t) + \sum_{\mu=1}^{\infty} \beta_\mu \sin(\mu\omega t)\,.$$

Mit der Entwicklung

$$\exp(-a \sin\omega t) = I_0(a) - 2 \sum_{\nu=0}^{\infty} (-1)^\nu I_{2\nu+1}(a) \sin[(2\nu+1)\omega t] + 2 \sum_{\nu=1}^{\infty} (-1)^\nu I_{2\nu}(a) \cos(2\nu\omega t) \tag{11}$$

erhält man:

$$\langle I_B \rangle = -q\, A\, N_t \langle c_p \rangle p_1 \left\{1 + \alpha_0 \left[1 + \frac{n_1}{n_{so}} I_0(a)\right] - \frac{n_1}{n_{so}} \sum_{\nu=1}^{\infty} (-1)^\nu I_{2\nu}(a)\, \alpha_{2\nu} + \frac{n_1}{n_{so}} \sum_{\nu=0}^{\infty} (-1)^\nu I_{2\nu+1}(a)\, \beta_{2\nu+1}\right\}$$

Im Falle kleiner Variationen gilt $a \ll 1$, und die modifizierten Besselfunktionen $I_\nu(a)$ können durch die asymptotischen Ausdrücke für kleine Argumente ersetzt werden:

$$I_\nu(a) = \frac{1}{\nu!}\left(\frac{a}{2}\right)^\nu \qquad \nu \geqslant 1$$

$$I_0(a) = 1 + \frac{a^2}{4}\,.$$

Bricht man die beiden Summen in der Beziehung für $\langle I_B \rangle$ mit a^2 ab, so ergibt sich:

$$\langle I_B \rangle = -q\, A\, N_t \langle c_p \rangle p_1 \left[1 - \alpha_0\left(1 + \frac{n_1}{n_{so}} + \frac{n_1}{n_{so}} \frac{a^2}{4}\right) + \alpha_2 \frac{n_1}{n_{so}} \frac{a^2}{8} + \beta_1 \frac{n_1}{n_{so}} \frac{a}{2}\right]. \tag{12}$$

Zur approximativen Berechnung der drei in (12) vorkommenden Fourierkoeffizienten wird (9) unter Weglassung der zeitabhängigen Terme im Koeffizienten von f_t vereinfacht zu:

$$\frac{df_t}{dt} = -f_t\left[\langle c_n\rangle n_{so}\left(1+\frac{a^2}{4}\right) + \langle c_p\rangle p_{so}\left(1+\frac{a^2}{4}\right) + \langle c_n\rangle n_1 + \langle c_p\rangle p_1\right]$$
$$+ \langle c_n\rangle n_{so}\exp(a\sin\omega t) + \langle c_p\rangle p_1 .$$

Mit der Entwicklung (11) bis zu quadratischen Termen in a erhält man:

$$\frac{df_t}{dt} = -f_t g_0 + g_1 + a g_2 \sin\omega t - \frac{a^2}{4} g_2 \cos 2\omega t$$

wo $g_0 = \langle c_n\rangle\left[n_{so}\left(1+\frac{a^2}{4}\right)+n_1\right] + \langle c_p\rangle\left[p_{so}\left(1+\frac{a^2}{4}\right)+p_1\right]$

$g_1 = \langle c_n\rangle n_{so}\left(1+\frac{a^2}{4}\right) + \langle c_p\rangle p_1$

$g_2 = \langle c_n\rangle n_{so}$.

Für die partikuläre Lösung resultiert:

$$f_t = \int_{-\infty}^{t} \exp[-g_0(t-t')]\left(g_1 + a\, g_2 \sin\omega t' - \frac{a^2}{4}\, g_2 \cos 2\omega t'\right) dt'$$

und daraus:

$$f_t = \frac{g_1}{g_0} + \frac{a\, g_2}{g_0^2+\omega^2}(g_0 \sin\omega t - \omega\cos\omega t)$$
$$- \frac{a^2}{4}\frac{g_2}{g_0^2+4\omega^2}(2\omega\sin 2\omega t + g_0\cos 2\omega t) .$$

Die gesuchten Fourierkoeffizienten ergeben sich zu:

$$\alpha_0 = \frac{g_1}{g_0} \qquad \alpha_2 = -\frac{a^2}{4}\frac{g_0 g_2}{g_0^2+4\omega^2} \qquad \beta_1 = a\frac{g_0 g_2}{g_0^2+\omega^2} .$$

Mit diesen Resultaten lässt sich (12) auswerten. Unter Weglassung des Terms, welcher a^4 enthält, wird:

$$\langle I_B\rangle = q\, A\, N_t \langle c_n\rangle\langle c_p\rangle n_i^2 \frac{a^2}{2}\frac{1}{g_0}\frac{\omega^2}{g_0^2+\omega^2} .$$

An dieser Stelle darf man nun a^2 gegen 1 vernachlässigen. So ergibt sich für den Mittelwert f_{to} der Fermifunktion:

$$f_{to} = \alpha_0 = \frac{g_1}{g_0} = \frac{1}{1+(p_{so}/p_1)} = \frac{1}{1+(n_1/n_{so})}$$

und damit:

$$g_0 = \frac{\langle c_n\rangle n_{so}}{f_{to}} + \frac{\langle c_p\rangle p_{so}}{1-f_{to}} = \frac{1}{\tau_{ns}} + \frac{1}{\tau_{ps}} = \frac{1}{\tau} .$$

Also wird schliesslich mit $a = q\Delta\psi_s/kT$:

$$\langle I_B \rangle = q\, A\, N_t \langle c_n \rangle \langle c_p \rangle n_i^2 \frac{\tau}{2} \left(\frac{q\,\Delta\psi_s}{kT} \right)^2 \frac{\omega^2 \tau^2}{1 + \omega^2 \tau^2} \,. \tag{13}$$

Sämtliche verwendeten MOS-Transistoren sind uns vom Centre Electronique Horloger in Neuchâtel zur Verfügung gestellt worden, wofür dem CEH und Herrn F. Leuenberger auch an dieser Stelle gedankt werde. Dem Schweizerischen Nationalfonds sind wir für die Unterstützung dieser Arbeit zu Dank verpflichtet.

LITERATURVERZEICHNIS

[1] J. S. Brugler and P. G. A. Jespers, *Charge Pumping in MOS Devices,* IEEE Trans. on El. Dev., ED-*16,* 297 (1969).

[2] A. G. Jordan and N. A. Jordan, *Theory of Noise in Metal Oxide Semiconductor Devices,* IEEE Trans. on El. Dev., ED-*12,* 148 (1965).

[3] K. H. Zaininger, *Semiconductor Surface Physics,* in: *Field Effect Transistors,* Edit. Wallmark and Johnson (Prentice Hall, 1966).

[4] E. H. Nicollian and A. Goetzberger, *The Si–SiO_2 Interface – Electrical Properties as Determined by the Metal-Insulator-Silicon Conductance Technique,* Bell System Technical Journal *46,* 1055 (1967).

[5] K. Lehovec and A. Slobodskoy, *Impedance of Semiconductor-Insulator-Metal Capacitors,* Solid-St. Electron. *7,* 59 (1964).

[6] W. Shockley and W. T. Read, Jr., *Statistics of the Recombinations of Holes and Electrons,* Phys. Rev. *87,* 835 (1952).

[7] A. Goetzberger and S. M. Sze, *Metal-Insulator-Semiconductor (MIS) Physics,* in: *Applied Solid State Science,* Vol. 1, Edit. Wolfe (Academic Press, 1969).

[8] C. N. Berglund, *Surface States at Steam-Grown Silicon-Silicon Dioxide Interfaces,* IEEE Trans. on El. Dev., ED-*13,* 701 (1966).

Erweiterte Untersuchung des differentiellen Charge Pumping-Effekts bei MOS-Transistoren

von **J. Golder**

Institut für angewandte Physik der Universität Basel

(17. 6. 1971)

Summary. As a consequence of variations of the gate potential, interface states in MOS-transistors are filled and emptied by capture and emission of both electrons and holes. If small periodical variations are applied, capture processes on the average slightly predominate, as was shown in a previous paper [1]. In p-enhancement-mode devices, where the contributions of holes and electrons can be observed separately, this predominance of capture-processes gives rise to a net DC-current 'pumped' from the p-regions (drain and source) into the substrate.

In the first sections, this paper deals with the influence of carrier diffusion through the substrate and the inversion layer. Theoretical and experimental results agree in showing that this influence is negligibly small in devices with low interface state densities. It is then shown, that the comparison of theory and experiment for such a device yields the values for the capture cross sections σ_n and σ_p (for electrons and holes respectively) as well as for the density of the interface states contributing to the effect.

1. Einleitung

Das Verhalten des MOS-Transistors bei kleinen Variationen der Gatespannung um einen Ruhewert wird in erster Näherung durch lineare Beziehungen beschrieben. Es ist deshalb möglich, die Änderungen der Ladungsträgerkonzentrationen in den Bändern sowie in Grenzschichtzuständen in der verbotenen Zone anhand eines Ersatzschemas zu diskutieren, dessen Elemente lediglich von den Gleichgewichtsgrössen abhängig sind.

In einer höheren Näherung stellt man einen endlichen gemittelten Substratstrom (Gleichstrom) als Funktion der variierenden Gatespannung (Wechselspannung) fest. Bei den zur vorliegenden und zu einer früheren Arbeit [1] durchgeführten Messungen war der gemittelte Substratstrom ungefähr 10^3- bis 10^4mal kleiner als die Amplitude des Wechselstroms, der durch die linearen Beziehungen beschrieben wird. Dieser Effekt wird als «Charge Pumping» bezeichnet.

Das von Brugler und Jespers [2] erstmals ausgeführte Charge Pumping-Experiment basiert nicht auf kleinen Variationen der Bandrandverbiegung, denn am Gate wird eine von Null ausgehende impulsförmige Spannung angelegt, deren Spitzenwert den MOS-Transistor in die Inversion zu bringen vermag. Der sich ergebende mittlere

Substratstrom $\langle I_B \rangle$ ist dann innerhalb der Messfehler proportional zur Repetitionsrate der normierten Impulse. Der Effekt kann also dahin gedeutet werden, dass jeder Gateimpuls eine normierte Ladung ins Substrat «pumpt». Die beim Experiment von Brugler und Jespers verwendeten Gatespannungsänderungen sind so gross, dass die Nichtlinearitäten im Zusammenhang zwischen Gatespannung und Bandrandverbiegung den Effekt wesentlich beeinflussen, wodurch die Analyse erschwert wird. Um diese Komplikation zu beseitigen, wurde in [1] und in der vorliegenden Arbeit eine differentielle Version des Charge Pumping-Effekts untersucht. In der theoretischen Behandlung kann man sich dann auf die Nichtlinearitäten der Einfangs- und Emissionsprozesse beschränken.

In [1] wurde das Umbesetzen von Grenzschichtzuständen untersucht. Für kleine periodische Variationen der Bandrandverbiegung um eine Ruhelage in schwacher Inversion ergab sich dabei aus der Shockley-Read-Hall-Statistik ein Überwiegen der Einfangsprozesse. Da alle Vorgänge sich dabei periodisch wiederholen, bleibt der Mittelwert der Ladung in den Grenzschichtzuständen konstant; demzufolge sind die Einfangsüberschüsse der Elektronen aus dem Leitungsband und der Löcher aus dem Valenzband gleich gross, so dass eine erzwungene Rekombination zustande kommt; es wird fortwährend positive Ladung vom Valenzband ins Leitungsband «gepumpt». Die Nachlieferung der rekombinierten Ladungsträger bedingt stetige Ströme von Elektronen und Löchern zur Oberfläche. In p-Enhancement-Mode MOS-Transistoren können diese beiden Beiträge getrennt beobachtet werden. Das n-Substrat liefert hierbei Elektronen an die Oberfläche, währenddem die p-Regionen (Drain und Source), als ohmsche Kontakte an die p-Inversionsschicht, den Löcheranteil übernehmen. Die Einfangsüberschüsse äussern sich bei der Messung am MOS-Transistor in von Null verschiedenen mittleren Strömen in den Zuleitungen zum Substrat bzw. zu den p-Regionen.

Bei der vorliegenden Untersuchung handelt es sich um eine Erweiterung der früheren Arbeit [1], indem der Einfluss der Diffusion der Elektronen vom Substratvolumen zur Oberfläche und der Löcher durch die schwache Inversionsschicht mitberücksichtigt wird. Es zeigt sich theoretisch und experimentell, dass der Einfluss der Diffusion bei Elementen mit geringer Zustandsdichte an der Grenzschicht vernachlässigbar klein ist. Messungen an derartigen Elementen dürfen demzufolge mit der in [1] hergeleiteten Theorie ausgewertet werden. Als Resultat lassen sich die Dichte N_t der am Effekt beteiligten Grenzschichtzustände sowie deren Einfangsquerschnitte σ_n und σ_p für Elektronen und Löcher angeben.

2. Die Ladungsträgerkonzentrationen unter Berücksichtigung der Diffusion

Für die Rechnung wird ein quaderförmiger MOS-Transistor mit der Gatefläche $A = w\,L$ angenommen. Dabei bedeuten w die Breite und L die Länge der Gate-Elektrode. Figur 1a zeigt einen Schnitt senkrecht zur Grenzschicht, welche in der Ebene $x = 0$ liegt. Die zum Umbesetzen der Grenzschichtzustände erforderlichen Ladungsträger werden durch einen Elektronenstrom I_n (parallel zur x-Richtung) von der Substratelektrode B und durch einen Löcherstrom I_p (parallel zur y-Richtung) von den p-Regionen (Drain D und Source S) zugeführt. Mit den Quasiferminiveaux

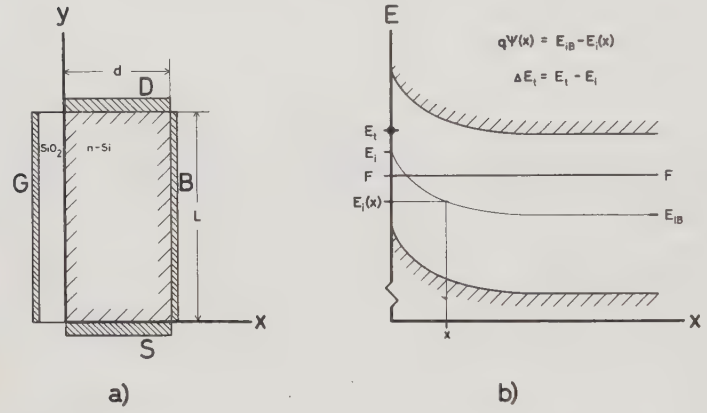
F_n und F_p für Elektronen bzw. Löcher lassen sich die Dichten der freien Ladungsträger wie folgt schreiben:

$$n = n_i \exp\left(\frac{F_n(x) - E_i(x)}{k\,T}\right) = n_B \exp\left(\frac{F_n(x) - F + q\,\Psi(x)}{k\,T}\right),$$

$$p = n_i \exp\left(-\frac{F_p(y) - E_i(x)}{k\,T}\right) = p_B \exp\left(-\frac{F_p(y) - F + q\,\Psi(x)}{k\,T}\right),$$

wobei $q\,\Psi(x) = E_{iB} - E_i(x)$ durch die Figur 1b erklärt wird. Ferner gilt für die Konzentrationen n_B und p_B im Substratvolumen wie in [1]:

$$n_B = n_i \exp\left(\frac{F - E_{iB}}{k\,T}\right), \quad n_B\,p_B = n_i^2 .$$



Figur 1

a) Schematischer Aufbau des MOS-Transistors: G = Gate-Elektrode; D = Drainanschluss (mit zugehöriger p-Zone); S = Sourceanschluss (mit zugehöriger p-Zone); B = Ohmscher Kontakt an das Substratmaterial.

b) Bänderschema für den Halbleiter in der Nähe der Grenzschicht.

Für die Stromdichten i_n und i_p erhält man unter Berücksichtigung der oben angegebenen Richtungsverhältnisse:

$$i_n = -\,q\,\mu_n\,n(x)\,\frac{\partial \Psi}{\partial x} + q\,D_n\,\frac{\partial n}{\partial x},$$

$$i_p = -\,q\,\mu_p\,p(x,y)\,\frac{\partial \Psi}{\partial x} - q\,D_p\,\frac{\partial p}{\partial y}.$$

Mit der Einsteinschen Beziehung zwischen der Diffusionskonstanten und der Beweglichkeit $D = (k\,T/q)\,\mu$ findet man nach Einsetzen der Konzentrationen n und p:

$$i_n = \mu_n\,n(x)\,\frac{\partial F_n(x)}{\partial x}, \tag{1}$$

$$i_p = \mu_p\,p(x,y)\,\frac{\partial F_p(y)}{\partial y}. \tag{2}$$

Für kleine Stromdichten i_p, i_n unterscheiden sich die Quasiferminiveaux nur wenig vom Ferminiveau, und die Konzentrationen lassen sich approximieren durch:

$$n = n_0 \left(1 + \frac{F_n - F}{k\,T}\right), \quad n_0 = n(F_n \equiv F)\,, \tag{3), (3a)}$$

$$p = p_0 \left(1 - \frac{F_p - F}{k\,T}\right), \quad p_0 = p(F_p \equiv F)\,, \quad n_0\,p_0 = n_i^2\,. \tag{4), (4a), (4b)}$$

Zunächst wird nun die Elektronenkonzentration $n(x = 0)$ an der Grenzschicht berechnet. Die Differentialgleichung (1) lässt sich separieren und integrieren, wenn man lediglich Terme von erster Ordnung in kleinen Grössen mitnimmt. Mit der Randbedingung $F_n(x = d) = F$ erhält man:

$$F_n(x = 0) - F = -\frac{i_n}{\mu_n}\frac{1}{n_i^2}\,P\,, \tag{5}$$

wobei für die totale Dichte P der Löcher pro Flächeneinheit gilt:

$$P = \int_0^d p_0(x)\,dx\,. \tag{6}$$

Einsetzen von (5) in (3) liefert mit dem gesamten Elektronenstrom $I_n = A\,i_n$ durch die Fläche A:

$$n(x = 0) = n_0(x = 0)\left(1 - \frac{q}{k\,T}\,I_n\,R_D\right),$$

$$R_D = \frac{P}{q\,A\,\mu_n\,n_i^2}\,. \tag{7}$$

Mit (4b) und (6) lässt sich (7) schreiben als:

$$R_D = \int_0^d \frac{dx}{q\,A\,\mu_n\,n_0(x)}\,.$$

Diese Beziehung zeigt, dass R_D den differentiellen Widerstand des Halbleiters für Elektronen zwischen dem Substratanschluss B bei $x = d$ und einem (fiktiven) Kontakt an der Grenzschicht ($x = 0$) darstellt. Wird nun der Strom I_n ausschliesslich durch die pro Zeiteinheit von Zuständen an der Grenzschicht ($x = 0$) aus dem Leitungsband eingefangenen Elektronen verursacht, so kann man mit der Nettoeinfangsrate U_{cn} schreiben: $I_n = q\,A\,U_{cn}$, und mit $n_0(x = 0) = n_s$ sowie $n(x = 0) = n_s^*$ ergibt sich:

$$n_s^* = n_s\left(1 - \frac{q^2 A\,U_{cn}\,R_D}{k\,T}\right). \tag{8}$$

Zur Berechnung von $F_p(y)$ wird zunächst mit (2) der totale Löcherstrom $I_p(y)$ durch einen Querschnitt senkrecht zur y-Achse bestimmt:

$$I_p(y) = w \int_0^d i_p \, dx = w \, \mu_p \frac{\partial F_p(y)}{\partial y} \int_0^d p_0(x) \, dx \, . \tag{9}$$

Die Grenzschichtzustände mit der Nettoeinfangsrate U_{cp} für Löcher seien homogen über die Oberfläche A verteilt. Der Löcherstrom I_p werde einzig durch den Einfang von Löchern aus dem Valenzband hervorgerufen. Zwischen $I_p(y + dy)$, $I_p(y)$ und der pro Zeiteinheit zwischen y und $y + dy$ eingefangenen Ladung $q \, U_{cp} \, w \, dy$ besteht demnach die folgende Beziehung:

$$I_p(y + dy) = I_p(y) - q \, U_{cp} \, w \, dy \, .$$

Als Randbedingung ergibt sich $I_p(y = L) = - I_p(y = 0)$, da beide p-Regionen (Drain D und Source S) gleichviel Löcher ins Substratmaterial liefern. Damit resultiert:

$$I_p(y) = q \, U_{cp} \, w \left(\frac{L}{2} - y \right) . \tag{10}$$

Schliesslich erhält man unter Berücksichtigung der Randbedingungen $F_p(y = 0) = F_p(y = L) = F$:

$$F_p(y) - F = \frac{q \, U_{cp}}{2 \, \mu_p \, P} (y \, L - y^2) \, .$$

Die einfache Theorie in [1] basiert auf einer von y unabhängigen Löcherkonzentration und damit auch auf einem von y unabhängigen Quasiferminiveau F_p. Die Berücksichtigung eines y-abhängigen Quasiferminiveaus führt auf das Ersatzschema mit verteilten Elementen von Figur 2a, welches im Anhang I näher beschrieben wird. Um die Rechnungen zu vereinfachen und an die Theorie von [1] anzuschliessen, führt man im folgenden näherungsweise ein von y unabhängiges Quasiferminiveau ein, welches dem linearen Mittelwert von $F_p(y)$ im Intervall $0 \leqslant y \leqslant L$ entspricht:

$$\langle F_p(y) - F \rangle = \frac{1}{L} \int_0^L [F_p(y) - F] \, dy \, ,$$

$$\langle F_p(y) - F \rangle = \frac{q^2 \, A \, U_{cp} \, R^*_{ch}}{2} \, , \tag{11}$$

$$R^*_{ch} = \frac{R_{ch}}{6} = \frac{L}{6 \, q \, \mu_p \, w \, P} \, . \tag{12}$$

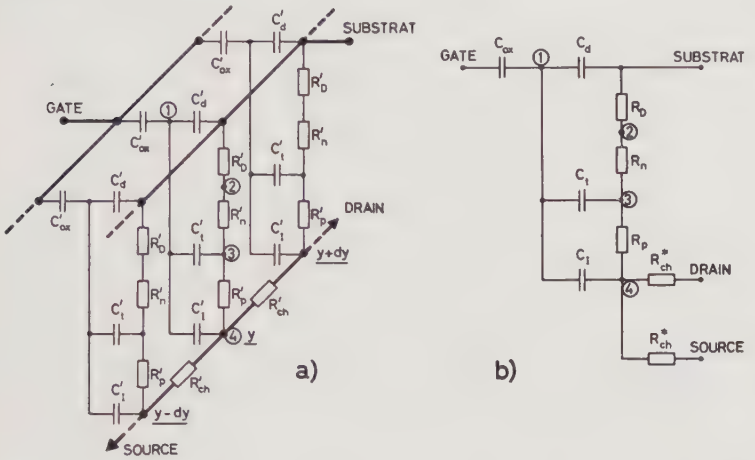
Aus (6) und (12) ist ersichtlich, dass R_{ch} den differentiellen Widerstand für Löcher zwischen Drain- und Sourcekontakt darstellt, und aus (10) bestimmt sich der von einer p-Region gelieferte Strom $I_p(y = 0)$ zu:

$$I_p(y = 0) = \frac{q \, A \, U_{cp}}{2} \, . \tag{13}$$

Einsetzen von (11) in (4) mit $p_0(x = 0) = p_s$ sowie $p(x = 0) = p_s^*$ liefert schliesslich:

$$p_s^* = p_s \left(1 - \frac{q^2 A U_{cp} R_{ch}^*}{2 k T}\right). \tag{14}$$

Durch die Einführung des y-unabhängigen Quasiferminiveaus F_p gelangt man zu dem einfacheren Ersatzschema von Figur 2b mit konzentrierten Elementen, welches ebenfalls im Anhang I näher erläutert wird.



Figur 2
a) Ersatzschema des MOS-Transistors, welches der y-Abhängigkeit der Löcherkonzentration Rechnung trägt. Wiedergegeben ist der Ausschnitt für $y - dy \leqslant y \leqslant y + dy$. Die eingekreisten Ziffern bezeichnen die im Text erwähnten Knoten.
b) Vereinfachtes Ersatzschema des MOS-Transistors. Die eingekreisten Ziffern bezeichnen die im Text erwähnten Knoten.

3. Modifikation der Theorie für kleine Variationen $\Delta\Psi_s$ der Bandrandverbiegung

Die zeitliche Änderung der Ladung $A\,Q_t$ in den Grenzschichtzuständen kann mit der Nettoeinfangsrate U_{cn} für Elektronen aus dem Leitungsband und der Nettoeinfangsrate U_{cp} für Löcher aus dem Valenzband folgendermassen geschrieben werden:

$$A \frac{dQ_t}{dt} = -\,q\,A\,U_{cn} + q\,A\,U_{cp}\,.$$

$A(dQ_t/dt)$ ist gleich dem Ladestrom I_t der Grenzschichtkapazität C_t in Figur 2b, $-\,q\,A\,U_{cn}$ und $q\,A\,U_{cp}$ sind die Beiträge des Leitungs- bzw. des Valenzbandes, welche in [1] mit I_B' und I_J' bezeichnet wurden:

$$I_B' = -\,q\,A\,U_{cn}\,, \quad I_J' = q\,A\,U_{cp}\,.$$

Unter Vernachlässigung der Diffusion ergeben sich für die Oberflächenkonzentrationen die Werte n_s und p_s, so dass man für U_{cn} und U_{cp} nach Shockley und Read [3] erhält:

$$U_{cn} = N_t \langle c_n \rangle [(1 - f_t)\, n_s - f_t\, n_1] , \tag{15a}$$

$$U_{cp} = N_t \langle c_p \rangle [f_t\, p_s - (1 - f_t)\, p_1] , \tag{15b}$$

wobei die einzelnen Grössen in [1] definiert wurden.

Im vorangehenden Kapitel ist gezeigt worden, dass die Berücksichtigung der Diffusion zu den in (8) und (14) gegebenen Oberflächenkonzentrationen n_s^* und p_s^* führt. Einsetzen von n_s^* und p_s^* in (15a) und (15b) liefert:

$$U_{cn}^* = \frac{N_t \langle c_n \rangle [(1 - f_t)\, n_s - f_t\, n_1]}{1 + N_t \langle c_n \rangle (1 - f_t) \dfrac{q^2}{k\,T} A\, R_D\, n_s} ,$$

$$U_{cp}^* = \frac{N_t \langle c_p \rangle [f_t\, p_s - (1 - f_t)\, p_1]}{1 + N_t \langle c_p \rangle f_t \dfrac{q^2}{2\,k\,T} A\, R_{ch}^*\, p_s} .$$

Für eine Korrektur erster Ordnung dürfen in den Nennern die Ruhewerte n_{s0}, p_{s0} und f_{t0} eingesetzt werden. Damit lassen sich die Raten U_{cn}^* und U_{cp}^* darstellen als:

$$U_n^* = N_t \langle c_n \rangle^* [(1 - f_t)\, n_s - f_t\, n_1] , \tag{16a}$$

$$U_{cp}^* = N_t \langle c_p \rangle^* [f_t\, p_s - (1 - f_t)\, p_1] , \tag{16b}$$

$$\langle c_n \rangle^* = \frac{\langle c_n \rangle}{1 + N_t \langle c_n \rangle (1 - f_{t0}) \dfrac{q^2}{k\,T} A\, R_D\, n_{s0}} ,$$

$$\langle c_p \rangle^* = \frac{\langle c_p \rangle}{1 + N_t \langle c_p \rangle f_{t0} \dfrac{q^2}{2\,k\,T} A\, R_{ch}^*\, p_{s0}} .$$

Der als Folge einer Variation $\Delta\Psi_s(t) = \Delta\Psi_s \sin\omega t$ entstehende mittlere Substratstrom $\langle I_B \rangle$ wurde in [1] (Gleichung (13)) angegeben. Umgeformt lautet diese Gleichung:

$$\langle I_B \rangle = q\, A\, \frac{n_i^2}{2} \left(\frac{q\, \Delta\Psi_s}{k\,T} \right)^2 \frac{\omega^2 \tau^2}{1 + \omega^2 \tau^2} \cdot \Phi \tag{17}$$

mit

$$\Phi = \left[\frac{n_1 + n_{s0}}{\langle c_p \rangle N_t} + \frac{p_1 + p_{s0}}{\langle c_n \rangle N_t} \right]^{-1} = \tau\, N_t \langle c_p \rangle \langle c_n \rangle . \tag{17a}$$

Der Einfluss der Diffusion kann nun durch Ersetzen von $\langle c_n \rangle$ durch $\langle c_n \rangle^*$ bzw. von $\langle c_p \rangle$ durch $\langle c_p \rangle^*$ berücksichtigt werden. Aus (17) und (17a) wird damit:

$$\langle I_B \rangle = q A \frac{n_i^2}{2} \left(\frac{q \Delta \Psi_s}{k T} \right)^2 \frac{(\omega \tau^*)^2}{1 + (\omega \tau^*)^2} \cdot \Phi^* , \tag{18}$$

$$\Phi^* = \left[\frac{n_1 + n_{s0}}{\langle c_p \rangle N_t} + \frac{p_1 + p_{s0}}{\langle c_n \rangle N_t} + \frac{q}{k T} n_i^2 A \left(\frac{R_{ch}^*}{2} + R_D \right) \right]^{-1}$$

$$= \tau^* N_t \langle c_p \rangle^* \langle c_n \rangle^* . \tag{18a}$$

Mit (7), (12) und mit (II.2) aus dem Anhang II lässt sich Φ^* schreiben als:

$$\Phi^* = \left[\frac{n_1 + n_{s0}}{\langle c_p \rangle N_t} + \frac{p_1 + p_{s0}}{\langle c_n \rangle N_t} + \frac{q}{k T} \left(\frac{\sqrt{2}\, L^2 n_{s0}}{6 \mu_p \lambda_{Dn}} + \frac{\lambda_{Dn} p_{s0}}{2\sqrt{2} \mu_n} \right) \right]^{-1}$$

$$= \tau^* N_t \langle c_p \rangle^* \langle c_n \rangle^* . \tag{19}$$

Wie zu erwarten, verschwindet der Einfluss der Diffusion bei kleinen Dichten N_t der Grenzschichtzustände: strebt N_t gegen Null, so geht Φ^* gegen Φ, wie aus (17a) und (19) hervorgeht.

Unter Berücksichtigung von $n_1 = n_i \exp(\Delta E_t / k T)$ und $n_{s0} = n_B \exp(q \Psi_{s0} / k T)$ sowie $n_1 p_1 = n_{s0} p_{s0} = n_i^2$ lässt sich zeigen, dass sowohl $\Phi(\Delta E_t, \Psi_{s0})$ als auch $\Phi^*(\Delta E_t, \Psi_{s0})$ ein absolutes Maximum aufweisen. Für den Fall beliebig kleiner N_t folgt unter der Voraussetzung energieunabhängiger $\langle c_n \rangle$ und $\langle c_p \rangle$ aus (17a):

$$\Delta E_t(\max) = \frac{k T}{2} \ln \frac{\langle c_p \rangle}{\langle c_n \rangle} , \tag{20a}$$

$$\Psi_{s0}(\max) = \frac{k T}{q} \left(\frac{1}{2} \ln \frac{\langle c_p \rangle}{\langle c_n \rangle} + \ln \frac{n_i}{n_B} \right) , \tag{20b}$$

$$\Phi(\max) = \frac{N_t}{4 n_i} \sqrt{\langle c_n \rangle \langle c_p \rangle} . \tag{20c}$$

Analog ergibt sich für endliche N_t aus (19):

$$\Delta E_t(\max) = \frac{k T}{2} \ln \frac{\langle c_p \rangle}{\langle c_n \rangle} , \tag{21a}$$

$$\Psi_{s0}^*(\max) = \frac{k T}{q} \left[\frac{1}{2} \ln \left(\frac{1}{\langle c_n \rangle N_t} + \frac{q}{k T} \frac{\lambda_{Dn}}{2\sqrt{2} \mu_n} \right) \right.$$

$$\left. - \frac{1}{2} \ln \left(\frac{1}{\langle c_p \rangle N_t} + \frac{q}{k T} \frac{\sqrt{2} L^2}{6 \mu_p \lambda_{Dn}} \right) + \ln \frac{n_i}{n_B} \right] , \tag{21b}$$

$$\Phi^*(\max) = \frac{1}{2 n_i} \left[\sqrt{\frac{1}{\langle c_n \rangle N_t} + \frac{q}{k T} \frac{\lambda_{Dn}}{2\sqrt{2} \mu_n}} \right.$$

$$\left. \times \sqrt{\frac{1}{\langle c_p \rangle N_t} + \frac{q}{k T} \frac{\sqrt{2} L^2}{6 \mu_p \lambda_{Dn}}} + \frac{1}{N_t \sqrt{\langle c_n \rangle \langle c_p \rangle}} \right]^{-1} . \tag{21c}$$

Der Term $(k\,T/q)\ln(n_i/n_B)$ in (20b) und (21b) ist gleich der Bandrandverbiegung, bei welcher die Mitte der verbotenen Zone an der Oberfläche mit dem Ferminiveau zusammenfällt ($E_i = F$). Bei den untersuchten Transistoren ergibt sich für diese Bandrandverbiegung ein Wert von –266 mV.

Ein Vergleich der Beziehungen (20b) und (20c) für verschwindende N_t mit den Resultaten (21b) und (21c) für endliche N_t zeigt, dass endliche N_t eine Verschiebung und wegen $\Phi^*(\max) < \Phi(\max)$ auch eine Erniedrigung des Maximums von $\langle I_B\rangle$ bewirken. In einiger Entfernung vom Maximum nimmt $\langle I_B\rangle$ in beiden Fällen als Funktion von Ψ_{s0} exponentiell ab, wie aus (17a) und (19) hervorgeht.

Die bisherige Analyse basiert auf der Annahme eines diskreten Energieniveaus E_t der Grenzschichtzustände. An der Grenzschicht Si-SiO_2 findet man jedoch kontinuierlich verteilte Zustände mit energieabhängiger Dichte $\bar{N}_t(E_t)$. Die starke Abnahme des Faktors Φ bzw. Φ^* als Funktion von E_t bewirkt jedoch eine Unterdrückung der Beiträge von Zuständen E_t, welche weiter als einige kT von $E_t(\max) = \Delta E_t(\max) + E_i$ entfernt liegen, vorausgesetzt, dass sowohl $N_t(E_t)$ als auch $\langle c_p\rangle$ und $\langle c_n\rangle$ als Funktionen des Arguments $|\,E_t - E_t(\max)\,|$ verglichen mit der Exponentialfunktion $\exp(|\,E_t - E_t(\max)\,|/k\,T)$ genügend schwach ansteigen. Für $\bar{N}_t(E_t)$ wird diese Voraussetzung gestützt durch Angaben von Nicollian und Goetzberger [4], nach denen $\bar{N}_t(E_t)$ in einem Bereich von 0,6 eV um die Mitte der verbotenen Zone um weniger als einen Faktor 2 variiert.

4. Experimentelle Ergebnisse

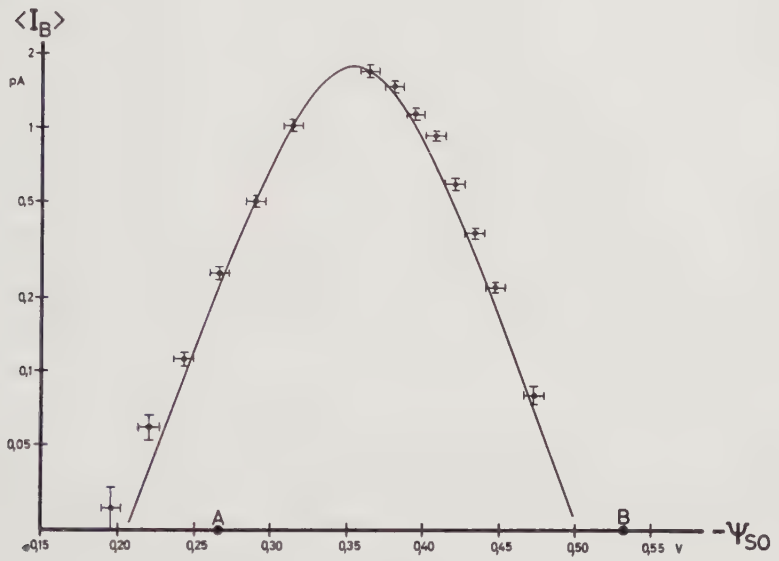
Zur Überprüfung der Abhängigkeit des gemittelten Substratstroms $\langle I_B\rangle$ von der Ruhelage Ψ_{s0} der Bandrandverbiegung können Messungen bei einer genügend hohen Frequenz ($\omega^2\,\tau^2 \gg 1$) herangezogen werden. Die einzige von Ψ_{s0} abhängige Grösse ist in diesem Fall Φ bzw. Φ^*. Mit dem in [1] beschriebenen Messaufbau wurden die in Figur 3 aufgetragenen Messpunkte an einem MOS-Transistor mit niedriger Dichte von Grenzschichtzuständen N_t bestimmt. Für diesen MOS-Transistor erwartet man, dass die einfache Theorie ohne Berücksichtigung der Diffusion in guter Näherung angewendet werden kann. Die ausgezogene Kurve wurde demzufolge auf Grund der Beziehungen (17) und (17a) berechnet. Da die Dichte N_t der Zustände an der Grenzschicht sowie die Einfangswahrscheinlichkeiten $\langle c_p\rangle$ und $\langle c_n\rangle$ nicht bekannt sind, muss die gerechnete Kurve den Messpunkten angepasst werden. Unter der Bedingung $\Delta E_t = \Delta E_t(\max)$ lässt sich die Beziehung (17) für $\omega^2\,\tau^2 \gg 1$ in der Form

$$\langle I_B\rangle = \frac{q\,A\,N_t}{16\,\tau(\max)}\left[\cosh\left(\frac{q\,\Psi_{s0} - q\,\Psi_{s0}(\max)}{k\,T}\right) + 1\right]^{-1}\left(\frac{q\,\Delta\Psi_s}{k\,T}\right)^2$$

schreiben. Logarithmisches Auftragen von $\langle I_B\rangle$ ermöglicht somit das Anpassen durch geeignetes Verschieben der gerechneten Kurve parallel zu beiden Achsen ohne Änderung der Kurvenform.

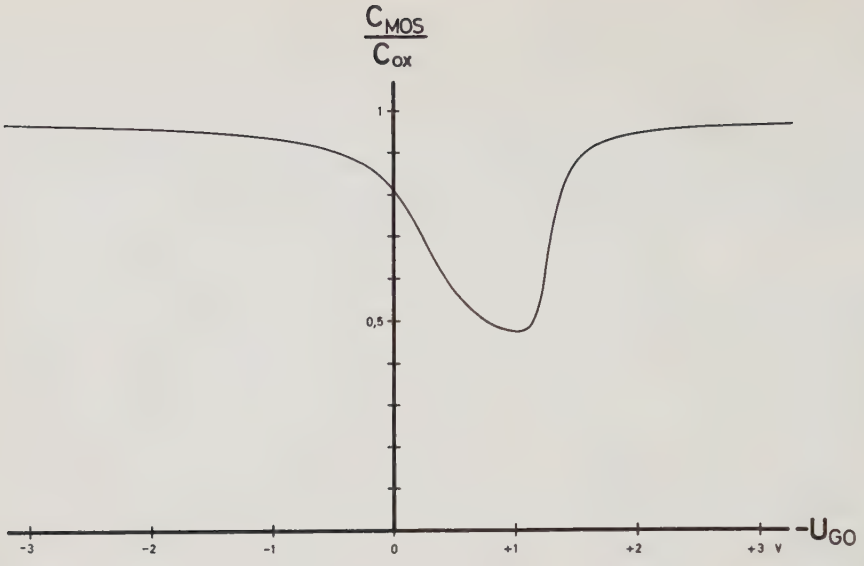
Die gute qualitative Übereinstimmung zwischen Messungen und der auf einem diskreten Niveau der Grenzschichtzustände basierenden Theorie zeigt, dass offensichtlich die Beiträge aller anderen Zustände genügend unterdrückt werden, so dass die

Annahmen im 3. Kapitel über die Energieabhängigkeit der Dichte $\bar{N}_t$ sowie der Einfangswahrscheinlichkeiten vernünftig erscheinen. Die Anpassung ergibt folgende Werte: $\Psi_{s0}(\max) = -364$ mV und $\langle I_B \rangle(\max) = 1{,}82$ pA. Zusätzlich wurde die zu $\Psi_{s0}(\max)$ gehörige Zeitkonstante $\tau(\max)$ zu 33 μs bestimmt. Unter Berücksichtigung von $\langle c \rangle = \sigma v_{th}$ (mit $v_{th} = 10^7$ cm · s^{-1}) folgen aus (20b) und (20c) die Einfangsquerschnitte: $\sigma_n = 1{,}5 \cdot 10^{-12}$ cm^2 und $\sigma_p = 1{,}6 \cdot 10^{-15}$ cm^2. Für die Dichte der Grenzschichtzustände, welche zum Effekt beitragen, ergibt sich $N_t = 1{,}3 \cdot 10^8$ cm^{-2}. Diese Resultate werden im nächsten Kapitel diskutiert.



Figur 3
Der mittlere Substratstrom $\langle I_B \rangle$ als Funktion der Ruhelage Ψ_{s0} der Bandrandverbiegung. Messdaten: $\Delta\Psi_s = 16$ mV, $f = 50$ kHz, $T = 27$ °C. Der Punkt A bezeichnet die Bandrandverbiegung, bei welcher die Oberfläche eigenleitend ist, der Punkt B den Einsatz der starken Inversion.

Der Zusammenhang zwischen Bandrandverbiegung Ψ_{s0} und Gatevorspannung U_{GO} wurde durch Integration aus der MOS-Kapazitätskurve von Figur 4 gewonnen. Dieses Verfahren liefert bei derart kleinen Elementen keine genauen Ergebnisse, da zusätzliche Kapazitäten (Überlappung der Gate-Elektrode mit den p-Regionen, Randfehler, Anschlussfläche für die Gatezuführung) mitgemessen werden. Damit resultiert eine zum grössten Teil von der Vorspannung U_{GO} unabhängige Verschiebung der Kurve nach oben. Die Gate-Kontaktierungsfläche bildet, obschon sie über einer bedeutend dickeren Oxidschicht liegt, eine parasitäre MOS-Kapazität, so dass die Kurve etwas verzerrt wird. Diese Schwierigkeit liesse sich durch Verwendung von Elementen mit bedeutend grösseren aktiven Gate-Flächen beheben. Gleichzeitig würde es möglich, mit derselben apparativen Auflösung den Effekt über ein grösseres Intervall der Bandrandverbiegung zu verfolgen. Damit könnte dann auch der Zusammenhang zwischen der Zeitkonstanten τ und der Bandrandverbiegung überprüft



Figur 4
Experimenteller Zusammenhang zwischen C_{MOS} und der Gate-Vorspannung U_{GO}. Die Streukapazitäten wurden vor der Normierung in Abzug gebracht.
Messdaten: $f = 1$ kHz, $\Delta U_G = 25 \mathrm{m} V_{eff}$.

werden. Die Zeitkonstante τ ergibt sich aus Messungen bei tiefen Frequenzen, welche mit sehr kleinen Amplituden der Gate-Wechselspannung durchgeführt werden müssen. Der Grund dafür liegt darin, dass bei tiefen Frequenzen Terme von höherer als zweiter Ordnung in $\Delta\Psi_s$ wichtiger werden, da ihre Frequenzabhängigkeit von derjenigen des Terms zweiter Ordnung verschieden ist. Diese Aussage wird durch die im Anhang von [1] angegebene Fourier-Entwicklung für die Fermifunktion f_t plausibel gemacht. Als Folge davon treten bei niedrigen Frequenzen bereits bei kleineren $\Delta\Psi_s$ Abweichungen von der Proportionalität $\langle I_B \rangle \sim \Delta\Psi_s^2$ auf als bei höheren Frequenzen ($\omega^2 \tau^2 \gg 1$). Sobald aber die Proportionalität $\langle I_B \rangle \sim \Delta\Psi_s^2$ nicht mehr erfüllt ist, stimmt auch der Frequenzgang nicht mehr mit dem in (17) angegebenen überein, und dadurch wird die Extraktion von τ verunmöglicht. Bei MOS-Transistoren mit kleiner Gate-Fläche erreicht man mit derart kleinen Werten von $\Delta\Psi_s$ rasch die Auflösungsgrenze des Elektrometers, so dass die zur Bestimmung von τ notwendigen Messungen nicht mehr durchgeführt werden können.

5. Diskussion

Aus der guten qualitativen Übereinstimmung der experimentellen und der theoretischen Ergebnisse darf geschlossen werden, dass die Vorstellung, welche der Analyse zugrunde liegt, den physikalischen Tatsachen gerecht wird.

Aus der Anpassung der theoretischen Kurve an die Messwerte von Figur 3 ergibt sich für die Dichte der Grenzschichtzustände $N_t = 1{,}3 \cdot 10^8 \mathrm{cm}^{-2}$. In dieser Zahl sind nur diejenigen Zustände enthalten, welche zum Effekt wirklich beitragen, das heisst,

welche in einer Umgebung von der Grössenordnung kT um $E_t(\max)$ liegen. Die Annahme einer energieunabhängigen Dichteverteilung $\bar{N}_t$ führt auf den Wert $\bar{N}_t \approx 3 \cdot 10^9\ \mathrm{cm}^{-2}\ (\mathrm{eV})^{-1}$, was nach Nicollian und Goetzberger [4] für eine Grenzschicht mit kleiner Zustandsdichte vernünftig ist.

Der Wert $\sigma_p = 1{,}6 \cdot 10^{-15}\ \mathrm{cm}^2$ für den Einfangsquerschnitt für Löcher stimmt grössenordnungsmässig ebenfalls mit den Angaben von Nicollian und Goetzberger [4] überein.

Aus der qualitativen Übereinstimmung von Theorie und Experiment und den sich ergebenden vernünftigen Werten für N_t und σ_p schliesst man, dass auch der Wert $\sigma_n \sim 1{,}5 \cdot 10^{-12}\ \mathrm{cm}^2$ für den Einfangsquerschnitt für Elektronen grössenordnungsmässig reell ist. In der Literatur fanden sich keine Angaben für σ_n von Zuständen in der unteren Hälfte der verbotenen Zone. Wie aus der von Many, Goldstein und Grover [5] veröffentlichten Zusammenstellung hervorgeht, sind Unterschiede von einigen Grössenordnungen zwischen σ_n und σ_p an Halbleiteroberflächen nicht ungewöhnlich. Genauere Angaben werden jedoch erst auf Grund einer exakteren Analyse möglich.

Zur quantitativen Überprüfung der Theorie ist es erforderlich, die Energieverteilung der Grenzschichtzustände sowie deren Einfangsquerschnitte zu bestimmen, was die genaue Kenntnis des Zusammenhangs zwischen der Bandrandverbiegung Ψ_s und der Gatevorspannung U_G voraussetzt. Auf die Schwierigkeiten, welche bei der Extraktion dieses Zusammenhangs bei kleinen MOS-Systemen auftreten, ist bereits im Kapitel 4 hingewiesen worden. Daher können erst Messungen an grossflächigeren MOS-Systemen die gewünschten quantitativen Zusammenhänge liefern.

Die theoretische Untersuchung ist vor allem in zwei Punkten unvollständig. Bei höheren Frequenzen, bei denen insbesondere die Ladungsträger der Inversionsschicht das momentane Gleichgewicht nicht mehr erreichen, wirken die Grenzschichtzustände als Erzeugungs- und Rekombinationszentren, welche Überschüsse der Ladungsträgerkonzentrationen ausgleichen. Ein Vorversuch mit Frequenzen bis zu einigen zehn Megahertz zeigte, dass der Frequenzgang im Bereich oberhalb 100 kHz wesentlich mit der Dichte der Grenzschichtzustände zusammenhängt.

Die zweite Unvollständigkeit besteht darin, dass der Einfluss der Diffusion nur näherungsweise berücksichtigt wurde. Dieses Vorgehen schien gerechtfertigt, da die Zusammenhänge an Transistoren mit geringen Zustandsdichten überprüft wurden. Wie Kapitel 3 zeigt, ist dann der Einfluss der Diffusion vernachlässigbar.

Bei grösseren Zustandsdichten können die Diffusionsströme derart starke Abweichungen vom momentanen Gleichgewicht erzeugen, dass bereits vor dem Erreichen der Grenzfrequenz der Inversionsschicht merkliche Rekombination und Erzeugung über die Grenzschichtzustände einsetzt. Experimentell könnte die Abhängigkeit des Effekts von der Zustandsdichte an der Grenzschicht mittels künstlicher Vermehrung der Grenzschichtzustände untersucht werden.

Anhang I

Lineare Ersatzschemata zur Analyse des Kleinsignalverhaltens

Die Analyse des dynamischen Verhaltens der Grenzschichtzustände basiert auf der Shockley-Read-Hall-Statistik, aus welcher sich die Einfangsraten U_{cn} und U_{cp}

für Elektronen aus dem Leitungsband sowie für Löcher aus dem Valenzband in der folgenden Form ergeben [3]:

$$U_{cn} = \langle c_n \rangle N_t \left[(1 - f_t)\, n_s - f_t\, n_1\right],$$

$$U_{cp} = \langle c_p \rangle N_t \left[f_t\, p_s - (1 - f_t)\, p_1\right],$$

wobei die Grössen in [1] definiert wurden.

Betrachtet man kleine Variationen der Konzentrationen n_s und p_s sowie der Fermifunktion f_t für Ladungsträger in den Grenzschichtzuständen um ein Gleichgewicht, so lässt sich schreiben:

$$n_s = n_B \exp\left(\frac{F_{ns} - F + q\,\Psi_s}{k\,T}\right) = n_{s0} + \delta n_s,$$

$$p_s = p_B \exp\left(-\frac{F_{ps} - F + q\,\Psi_s}{k\,T}\right) = p_{s0} + \delta p_s,$$

$$f_t = \left[1 + \exp\left(\frac{\Delta E_t + E_{iB} - q\,\Psi_s - F_t}{k\,T}\right)\right]^{-1} = f_{t0} + \delta f_t,$$

n_{s0}, p_{s0}, f_{t0} bedeuten dabei die Gleichgewichtswerte.

Mit den Spannungen

$$v_n = -\frac{1}{q}(F_{ns} - F), \tag{I.1}$$

$$v_p = -\frac{1}{q}(F_{ps} - F), \tag{I.2}$$

$$v_t = -\frac{1}{q}(F_t - F), \tag{I.3}$$

erhält man näherungsweise:

$$\delta n_s = \frac{q}{k\,T}\, n_{s0}(\Delta\Psi_s - v_n), \tag{I.4}$$

$$\delta p_s = \frac{q}{k\,T}\, p_{s0}(v_p - \Delta\Psi_s), \tag{I.5}$$

$$\delta f_t = \frac{q}{k\,T}\, f_{t0}(1 - f_{t0})\,(\Delta\Psi_s - v_t). \tag{I.6}$$

Berücksichtigt man, dass die Einfangsraten U_{cn} und U_{cp} im Gleichgewicht einzeln verschwinden, so resultieren die Bedingungen:

$$(1 - f_{t0})\, n_{s0} = f_{t0}\, n_1,$$

$$f_{t0}\, p_{s0} = (1 - f_{t0})\, p_1.$$

Damit ergeben sich die Einfangsraten zu:

$$U_{cn} = \langle c_n \rangle N_t \frac{q}{k\,T} n_{s0}(1 - f_{t0})\,(v_t - v_n)\,,$$

$$U_{cp} = \langle c_p \rangle N_t \frac{q}{k\,T} p_{s0} f_{t0}(v_p - v_t)\,.$$

Im Depletion- und Inversionsgebiet werden an der Grenzschicht nur in sehr geringem Masse Majoritätsladungsträger gespeichert. Der gesamte Majoritätsladungsträgerstrom i_n nach der Grenzschicht wird deshalb durch die Einfangsprozesse bestimmt. Minoritätsladungsträger werden teils eingefangen und teils in der Inversionsschicht gespeichert. Damit lassen sich die Bilanzgleichungen für die Ladungsträger im Leitungsband (I.7), in den Grenzschichtzuständen (I.8) sowie im Valenzband (I.9) für ein Flächenelement der Breite w und der Länge dy am Ort y aufstellen:

$$(q\,U_{cn} - i_n)\,w\,dy = 0\,, \tag{I.7}$$

$$(-\,q\,U_{cn} + q\,U_{cp})\,w\,dy = \frac{d}{dt}(\delta Q_t)\,, \tag{I.8}$$

$$-\,q\,U_{cp}\,w\,dy + I_p(y) - I_p(y + dy) = \frac{d}{dt}(\delta Q_p)\,. \tag{I.9}$$

Aus (5) und (7) folgt mit $F_n(x = 0) = F_{sn}$ sowie mit (I.1):

$$i_n = \frac{v_n}{A\,R_D}\,.$$

Aus (9), (12) und (I.2) ergibt sich:

$$I_p(y) - I_p(y + dy) = \frac{L}{R_{ch}\,dy}[v_p(y - dy) - v_p(y)] + \frac{L}{R_{ch}\,dy}[v_p(y + dy) - v_p(y)]$$

$$= \frac{1}{R'_{ch}}[v_p(y - dy) - v_p(y)] + \frac{1}{R'_{ch}}[v_p(y + dy) - v_p(y)]\,.$$

Für die Ladungen δQ_p und δQ_t gelten die Beziehungen: $\delta Q_p = q\,P\,w\,dy$, $\delta Q_t = -\,q\,N_t f_t\,w\,dy$. Daraus folgt mit (I.2) und (I.3):

$$\frac{d}{dt}(\delta Q_p) = \frac{q^2}{k\,T}\,w\,dy\,\frac{\lambda_{Dn}}{2\sqrt{2}}\,p_{s0}\left(\frac{dv_p}{dt} - \frac{d\Delta\Psi_s}{dt}\right) = C'_I\left(\frac{dv_p}{dt} - \frac{d\Delta\Psi_s}{dt}\right), \tag{I.10}$$

$$\frac{d}{dt}(\delta Q_t) = \frac{q^2}{k\,T}\,w\,dy\,N_t\,f_{t0}(1 - f_{t0})\left(\frac{dv_t}{dt} - \frac{d\Delta\Psi_s}{dt}\right) = C'_t\left(\frac{dv_t}{dt} - \frac{d\Delta\Psi_s}{dt}\right). \tag{I.11}$$

Für die totale Ladung δQ_s im Flächenelement $w\,dy$ gilt: $\delta Q_s = \delta Q_p + \delta Q_t + \delta Q_d$. δQ_d kommt durch die Verdrängung von Majoritätsladungsträgern von der Grenzschicht zustande. Die Auswertung der Maxwellschen Gleichung $\mathrm{div}\,\vec{D} = \varrho$ mit Hilfe des Gaußschen Satzes ergibt für die betrachtete Geometrie:

$$\frac{\varepsilon_0\,\varepsilon_{ox}}{d_{ox}}(\Psi_s - U_G) = \frac{\delta Q_s}{w\,dy}$$

(ε_{ox} = Dielektrizitätskonstante, d_{ox} = Dicke des Oxids) und daraus:

$$\frac{d}{dt}(\delta Q_s) = \frac{\varepsilon_0\,\varepsilon_{ox}}{d_{ox}}\,w\,dy\left(\frac{d\Delta\Psi_s}{dt} - \frac{d\Delta U_G}{dt}\right), \tag{I.12}$$

und weil das Substratvolumen als feldfrei angenommen wird, gilt weiter:

$$\frac{d}{dt}(\delta Q_d) = -\,C'_d\,\frac{d\Delta\Psi_s}{dt}.$$

Daraus resultieren die Knotengleichungen für das Ersatzschema in Figur 2a:

Knoten 1: $$C'_{ox}\left(\frac{d\Delta U_G}{dt} - \frac{d\Delta\Psi_s}{dt}\right) + C'_I\left(\frac{dv_p(y)}{dt} - \frac{d\Delta\Psi_s}{dt}\right) + C'_t\left(\frac{dv_t}{dt} - \frac{d\Delta\Psi_s}{dt}\right) - C'_d\,\frac{d\Delta\Psi_s}{dt} = 0,$$

Knoten 2: $$\frac{v_t - v_n}{R'_n} - \frac{v_n}{R'_D} = 0,$$

Knoten 3: $$\frac{v_n - v_t}{R'_n} + \frac{v_p(y) - v_t}{R'_p} + C'_t\left(\frac{d\Delta\Psi_s}{dt} - \frac{dv_t}{dt}\right) = 0,$$

Knoten 4: $$\frac{v_t - v_p(y)}{R'_p} + \frac{v_p(y+dy) - v_p(y)}{R'_{ch}} + \frac{v_p(y-dy) - v_p(y)}{R'_{ch}} + C'_I\left(\frac{d\Delta\Psi_s}{dt} - \frac{dv_p(y)}{dt}\right) = 0.$$

Die Knotenspannungen $u(K)$ ergeben sich zu $u(1) = \Delta\Psi_s$, $u(2) = v_n$, $u(3) = v_t$, $u(4) = v_p(y)$, und für die Elemente erhält man:

$$R'_n = \frac{k\,T}{\langle c_n\rangle\,N_t\,q^2\,n_{s0}(1 - f_{t0})\,w\,dy} \qquad C'_I = \frac{q^2}{k\,T}\,\frac{\lambda_{Dn}}{2\sqrt{2}}\,p_{s0}\,w\,dy,$$

$$R'_p = \frac{k\,T}{\langle c_p\rangle\,N_t\,q^2\,p_{s0}\,f_{t0}\,w\,dy} \qquad C'_t = \frac{q^2}{k\,T}\,N_t\,f_{t0}(1 - f_{t0})\,w\,dy,$$

$$R'_{ch} = R_{ch}\,\frac{dy}{L} \qquad C'_{ox} = \varepsilon_o\,\varepsilon_{ox}\,\frac{w\,dy}{d_{ox}},$$

$$R'_D = R_D\,\frac{L}{dy}.$$

Für das vereinfachte Ersatzschema von Figur 2b wird die gesamte Fläche $A = w\,L$ anstelle des Flächenelements $w\,dy$ betrachtet, und für das Quasiferminiveau F_p führt man den in (11) angegebenen Mittelwert ein. Analog zu den Beziehungen (I.7), (I.8) und (I.9) erhält man dann:

$$(q\,U_{cn} - i_n)\,A = 0,$$

$$(-\,q\,U_{cn} + q\,U_{cp})\,A = \frac{dQ_t}{dt},$$

$$-\,q\,U_{cp}\,A + 2\,I_p = -\,\frac{dQ_p}{dt}.$$

Aus (11) und (13) folgt mit (I.2):

$$I_p = I_p(y=0) = \frac{\langle F_p - F \rangle}{q} \frac{1}{R^*_{ch}} = -\frac{\langle v_p \rangle}{R^*_{ch}} .$$

Mit $Q_p = q\,A\,P$ und $Q_t = -q\,N_t\,A\,f_t$ wird analog zu (I.10) und (I.11):

$$\frac{dQ_p}{dt} = \frac{q^2}{k\,T} A \frac{\lambda_{Dn}}{2\sqrt{2}} p_{s0} \left(\frac{d\langle v_p \rangle}{dt} - \frac{d\Delta\Psi_s}{dt} \right) = C_I \left(\frac{d\langle v_p \rangle}{dt} - \frac{d\Delta\Psi_s}{dt} \right),$$

$$\frac{dQ_t}{dt} = \frac{q^2}{k\,T} A\,N_t f_{t0}(1-f_{t0}) \left(\frac{dv_t}{dt} - \frac{d\Delta\Psi_s}{dt} \right) = C_t \left(\frac{dv_t}{dt} - \frac{d\Delta\Psi_s}{dt} \right).$$

Die totale Ladung Q_s erfüllt die Beziehung:

$$Q_s = Q_p + Q_t + Q_d .$$

Entsprechend (I.12) wird:

$$\frac{dQ_s}{dt} = \frac{\varepsilon_0\,\varepsilon_{ox}}{d_{ox}} A \left(\frac{d\Delta\Psi_s}{dt} - \frac{d\Delta U_G}{dt} \right) = C_{ox} \left(\frac{d\Delta\Psi_s}{dt} - \frac{d\Delta U_G}{dt} \right)$$

und mit $(dQ_d/dt) = -C_d(d\Delta\Psi_s/dt)$ resultieren die Knotengleichungen für das Ersatzschema in Figur 2b:

$$\text{Knoten 1:} \quad C_{ox} \left(\frac{d\Delta U_G}{dt} - \frac{d\Delta\Psi_s}{dt} \right) + C_I \left(\frac{d\langle v_p \rangle}{dt} - \frac{d\Delta\Psi_s}{dt} \right) + C_t \left(\frac{dv_t}{dt} - \frac{d\Delta\Psi_s}{dt} \right) - C_d \frac{d\Delta\Psi_s}{dt} = 0 ,$$

$$\text{Knoten 2:} \quad \frac{v_t - v_n}{R_n} - \frac{v_n}{R_D} = 0 ,$$

$$\text{Knoten 3:} \quad \frac{v_n - v_t}{R_n} + \frac{\langle v_p \rangle - v_t}{R_p} + C_t \left(\frac{d\Delta\Psi_s}{dt} - \frac{dv_t}{dt} \right) = 0 ,$$

$$\text{Knoten 4:} \quad \frac{v_t - \langle v_p \rangle}{R_p} - \frac{2\langle v_p \rangle}{R^*_{ch}} + C_I \left(\frac{d\Delta\Psi_s}{dt} - \frac{d\langle v_p \rangle}{dt} \right) = 0 ,$$

mit den Knotenspannungen $u(1) = \Delta\Psi_s$, $u(2) = v_n$, $u(3) = v_t$, $u(4) = \langle v_p \rangle$ und den Elementen:

$$R_n = \frac{k\,T}{\langle c_n \rangle N_t\,q^2\,n_{s0}(1-f_{t0})\,A} \qquad C_I = \frac{q^2}{k\,T} A \frac{\lambda_{Dn}}{2\sqrt{2}} p_{s0} ,$$

$$R_p = \frac{k\,T}{\langle c_p \rangle N_t\,q^2\,p_{s0}\,f_{t0}\,A} \qquad C_t = \frac{q^2}{k\,T} A\,f_{t0}(1-f_{t0})\,N_t ,$$

$$R^*_{ch} = \frac{R_{ch}}{6} = \frac{1}{6} \frac{L}{q\,\mu_p\,w\,P} \qquad C_{ox} = \varepsilon_0\,\varepsilon_{ox} \frac{A}{d_{ox}} ,$$

$$R_D = \frac{P}{q\,\mu_n\,A\,n_i^2} .$$

Aus den Knotenspannungen in den Ersatzschemata lassen sich unter Benützung der Beziehungen (I.4), (I.5) und (I.6) die Variationen der Ladungsträgerkonzentrationen n_s und p_s sowie der Fermifunktion f_t für die Ladungsträger in den Grenzschichtzuständen direkt berechnen. Zu den in [1] beschriebenen Elementen wurden neu die Ersatzelemente für die Diffusion der Elektronen vom Substratvolumen zur Oberfläche (R'_D, R_D) sowie für die Diffusion der Löcher parallel zur Grenzschicht (R'_{ch}, R^*_{ch}) hinzugenommen.

Anhang II

Berechnung der totalen Dichte P der Löcher pro Flächeneinheit

Das Integral in Gleichung (6) lässt sich in zwei Teile aufspalten:

$$P = \int_0^d p_0(x)\, dx = \Delta P + p_B\, d\,.$$

ΔP bedeutet dabei die Überschussdichte der Löcher pro Flächeneinheit der Gate-Elektrode. Anhand der von Young [6] numerisch berechneten Kurven lässt sich für ΔP im Intervall $p_B < p_0(x = 0) \leqslant n_B$ (Flat-Band bis und mit schwache Inversion) die folgende Approximation angeben:

$$\Delta P = \frac{\lambda_{Dn}}{2\sqrt{2}}\, p_0(x = 0)\,,$$

$$\lambda_{Dn} = \sqrt{\frac{\varepsilon_0\, \varepsilon\, k\, T}{q^2\, n_B}} = \text{Debye-Länge im } n\text{-Substrat.}$$

Im Gleichgewichtszustand wird $p_0(x = 0) = p_{s0}$, und man erhält für die totale Dichte P der Löcher pro Flächeneinheit:

$$P = \frac{\lambda_{Dn}}{2\sqrt{2}}\, p_{s0} + p_B\, d\,. \tag{II.1}$$

Bei den untersuchten Transistoren mit $n_B = 5 \cdot 10^{14}\,\text{cm}^{-3}$ und $d = 2 \cdot 10^{-2}\,\text{cm}$ kann der zweite Term in (II.1) in guter Näherung vernachlässigt werden, sobald $p_{s0} \geqslant n_i$ gilt. Damit resultiert:

$$P \approx \frac{\lambda_{Dn}}{2\sqrt{2}}\, p_{s0}\,. \tag{II.2}$$

Meinem verstorbenen Lehrer, Herrn Prof. Dr. E. Baldinger, schulde ich Dank für seine wertvolle Anleitung zu dieser Arbeit und für die vielen Diskussionen darüber, Herrn PD Dr. W. Czaja für seine Anregungen und die kritische Durchsicht des Manuskripts. Ferner möchte ich Herrn F. Leuenberger vom Centre Electronique Horloger in Neuenburg für die MOS-Transistoren danken, die er stets zur Verfügung gestellt hat, sowie dem Schweizerischen Nationalfonds für die finanzielle Unterstützung dieser Untersuchung.

LITERATUR

[1] J. Golder und E. Baldinger, *Untersuchung des Zusammenhangs zwischen Grenzschichtzuständen und Charge Pumping-Effekt bei MOS-Transistoren*, Helv. phys. Acta *44*, 387 (1971).

[2] J. S. Brugler und P. G. A. Jespers, *Charge Pumping in MOS Devices*, IEEE Trans. on El. Dev. ED-*16*, 297 (1969).

[3] W. Shockley und W. T. Read, jr., *Statistics of the Recombinations of Holes and Electrons*, Phys. Rev. *87*, 835 (1952).

[4] E. H. Nicollian und A. Goetzberger, *The Si-SiO_2 Interface – Electrical Properties as Determined by the Metal-Insulator-Silicon Conductance Technique*, Bell System Technical Journal *46*, 1055 (1967).

[5] A. Many, Y. Goldstein und N. B. Grover, *Semiconductor Surfaces* (North-Holland Publishing, 1965), p. 390.

[6] C. E. Young, *Extended Curves of the Space Charge, Electric Field, and Free Carrier Concentration at the Surface of a Semiconductor, and Curves of the Electrostatic Potential Inside a Semiconductor*, J. appl. Phys. *32*, 329 (1961).

Curriculum vitae

Am 9. Dezember 1942 wurde ich, JOHANNES GOLDER, Bürger von Herdern TG, in Frauenfeld geboren als Sohn des Dr. med. Otto Golder, prakt. Arzt, seit 1949 in Ostermundigen BE, und der Dora Emma, geb. Wichtermann.

Nach dem Besuch der Primarschule in Frauenfeld und Ostermundigen trat ich im Frühjahr 1952 ins Städtische Progymnasium Bern und im Frühjahr 1956 in die Literarschule des Städtischen Gymnasiums Bern ein, wo ich im Herbst 1960 mit der Matura Typus B abschloss. Darauf begann ich an der Abteilung für Elektrotechnik der ETH Zürich meine Studien, die ich im Dezember 1964 mit dem Diplom als Elektroingenieur beendete. Anschliessend war ich als Entwicklungsingenieur in der Industrie tätig. Im Herbst 1966 immatrikulierte ich mich an der Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel, und gleichzeitig begann ich unter der Leitung von Herrn Prof. Dr. E. Baldinger, Vorsteher des Instituts für Angewandte Physik, mit der Ausarbeitung meiner Dissertation. Im Dezember 1968 legte ich bei Herrn Prof. K. Alder das Nebenfachexamen in Theoretischer Physik ab. Ende 1970, nach dem Tod von Herrn Prof. Dr. E. Baldinger, übernahm Herr PD Dr. W. Czaja die Betreuung meiner Arbeit.

An der ETH Zürich habe ich die Vorlesungen, Übungen und Praktika der Herren Professoren und Dozenten E. Baumann, K. Berger, J. P. Blaser, E. Böhler, F. Borgnis, E. Brandenberger, G. Calgari, W. Daenzer, C. Dubois, A. Dutoit, B. Eckmann, G. Epprecht, H. Gerber, W. Gerber, E. Gerecke, W. Guggenbühl, F. Held, W. Hug, H. Kern, F. Kummer, H. Leuthold, H. Lüthy, J. Müller, M. Müller, H. Rutishauser, J. R. de Salis, W. Saxer, E. Stiefel, M. J. O. Strutt, W. Traupel, A. Vogt, H. Weber, M. Wildi, H. Ziegler besucht.

In Basel habe ich die Vorlesungen und Übungen der Herren Professoren K. Alder, E. Baldinger, P. Diehl, P. Scherrer, H. R. Striebel, R. Wagner und der Privatdozenten W. Czaja und H. Rudin besucht.

Johannes Golder

Reinach, den 30. März 1971